A Model for Colour Naming and Comparing based on Conceptual Neighbourhood. An Application for Comparing Art Compositions

Zoe Falomir[a,*], Lledó Museros[b], Luis Gonzalez-Abril[c]

[a]*Cognitive Systems Dep., Spatial Cognition Research Centre, Universität Bremen, Enrique-Schmidt-Str. 5, 28359 Bremen, Germany*
[b]*Engineering and Computer Science Dep., Universitat Jaume I, Av. Vicent Sos Baynat s/n, E-12071, Castellón, Spain*
[c]*Applied Economics I Dep., Universidad de Sevilla, Avda Ramon y Cajal, 1, Sevilla, E-41018, Spain*

Abstract

A computational model for Qualitative Colour Description, named the QCD model, is defined using the Hue, Saturation and Luminance colour space. This model can name *rainbow* colours, *pale*, *light* and *dark* colours, and colours in the *grey* scale, and it has been parameterised by participants of a study in two universities in Spain: *University Jaume I* and *University of Sevilla*. The relational structure of the QCD model is analysed by means of a conceptual neighbourhood diagram and it is used to formulate a measure of similarity for solving absolute and relative comparisons of qualitative colours. Moreover, a similarity measure between colour compositions, called *SimQCDI*, is also developed. A survey test on several art compositions is carried out and the results obtained by the participants are analysed and compared to the computational results provided by the *SimQCDI*. Also, a comparison to the standard RGB Colour Histogram similarity method is carried out, which shows that the proposed similarity is more intuitive and that the results obtained are similar with respect to quantification. Finally, the cognitive adequacy of the QCD model is also analysed.

Keywords: Qualitative Representations, Colour Model, Colour Naming, Similarity Measure, Complementary Colours, Conceptual Neighbourhood Diagrams, Image Similarity, Cognitive Adequacy

1. Introduction

Human beings can see coloured surfaces because the light emitted by luminous objects, such as the sun or light bulbs, is reflected by these surfaces into their eyes and a proper nervous system allow them to experience it. There may be a *light* independent of an observer, but there is no colour independent of an observer, because *colour* is a psychological phenomenon that arises only within an observer [1].

Human beings are called trichromats due to their three types of cone cells, or photoreceptors, that can capture three different light wavelengths (short, medium and long) and any colour can

*Corresponding author. Z. Falomir is at Cognitive Systems (CoSy) Department, Spatial Cognition Centre, Universität Bremen, Enrique-Schmidt-Str. 5, 28359 Bremen, Germany.

Email addresses: zfalomir@informatik.uni-bremen.de (Zoe Falomir), museros@uji.es (Lledó Museros), luisgon@us.es (Luis Gonzalez-Abril)

be matched with some mixture of three others [1]. That is, coloured objects can be observed as different because their surfaces reflect different proportions of light at different wavelengths. In fact, people can distinguish a palette of around 7,000,000 colours [2]. However, in the practice, people communicating in English language seem to get along well with no more than about a dozen colour names.

Furthermore, a real fact in human cognition is that people go beyond the purely perceptual experience to classify things as members of categories and attach linguistic labels to them, and colour is no exception. For example, fresh blood and ripe tomatoes are both classified as *red*, even though these objects reflect different wavelengths [1]. Humans also attach colours to objects and think about them qualitatively and as a constant: *white* wine, *blue* sea, etc. even though they know that *white* wine is in fact *yellowish* or *golden* and that the sea is sometimes *grey* or *turquoise*. In fact, some studies concluded that the basic colours that can be named by people are limited to about 10-20 [3].

Human beings are not aware of how wavelengths are perceived by the photoreceptors of their eyes. What they are conscious of, is that they describe and compare colours by their names, that is, qualitatively, and viceversa. Colour representations in the mind are activated when colour words are read or heard [4]. Other studies on representing object colour in language comprehension concluded that object colour is represented differently to other object properties such as shape and orientation [5]. And experimentation results showed that coherent colour representation of objects enhances people's object recognition and conceptualization [7].

A computational approach for colour-naming can be easily interpreted by human users and then used for enhancing user-machine communication in many applications. A qualitative colour description can be assigned a meaning by relating it to an ontology [6] and, in this way, it could be interpretable by intelligent web agents and also by robotic agents. Thus, how colours are labelled is important because naming involves conceptual alignment with human cognition, meaning and human understanding.

Given that there are no experimental results demonstrating the higher consistency with human perception of a colour space over any other, this approach, which deals with the challenge of defining a computational model for cognitive and adaptive colour-naming, has chosen the Hue, Saturation and Luminance (HSL) colour space as a baseline since, according to since, according to Clark [8], it captures the entire gamut[1] of colours that humans can perceive.

Another challenge appears when trying to compare two colour names. How is it possible to define the degree of similarity between *blue* and *purple* colours? Or which colour is darker: *grey* or *dark blue*? Or which colour is more yellowish: *orange* or *pink*? According to Palmer [1] human beings have a relational structure of colours in the mind: '*Without a relational structure we would not experience different colours as being more closely related to each other (...) Nor would we experience grey as being intermediate between white and black; we would experience them only as different'*. Therefore, to be able to compare colour names cognitively, they must be organised in a colour space. The model for colour naming and comparing defined in this paper is based on the relational structure or conceptual neighbourhood of colours in the HSL space.

The rest of the paper is organised as follows. Section 2 presents related work on colour naming and comparing. The model for QCD is presented in Section 3 and parameterised in Section 4. Section 5 explains the relational structure of the QCD model using a conceptual neighbourhood diagram which is used in Section 6 to define a colour similarity measure to solve absolute

[1] A colour gamut is the area enclosed by a colour space in three dimensions.

and relative comparisons. Section 8 explains how to obtain the complementary of a given colour name in the QCD model. Section 9 describes how to compare two compositions/images using colour similarity (*SimQCDI*). Section 10 shows some experimentation carried out considering a scenario of art compositions and using *SimQCDI* to calculate the similarity between them. Section 11 compares the similarity values and the results obtained by a survey carried out to 109 participants. Then the cognitive adequacy of the QCD model is discussed and the similarity between qualitative colours (*SimQCD*) is analysed with respect to the literature. In order to study if the image similarity obtained by the QCD model (*SimQCDI*) is more intuitive or consistent with human perception than standard colour-based image descriptors such as RGB histograms, a comparative is carried out in this section. Finally, conclusions and future work are explained.

2. Related Work on Colour Naming and Comparing

From the point of view of colour vision psychophysics and colour categorization, colour models can be classified as: (a) descriptive or topological models, (b) geometric models, and (c) models based on chromaticity diagrams. Descriptive colour appearance models represent three subjective dimensions of colour and variation of them in *topological* terms defining spaces, such as: RGB (Red, Green and Blue), HSL (Hue, Saturation and Luminance), HSV/HSB (Hue, Saturation and Value or Brightness) and HSI (Hue, Saturation and Intensity). Some colour appearance models fulfill *geometrical* assumptions, i.e. the Munsell colour solid [9] where the perceptual distance between two colours is measured by the number of just noticeable differences [10]. Colour models based on chromaticity diagrams are derived from a mixture of *physical* characteristics of three ideal light sources (red, green and blue) and they are defined *mathematically* as radial basis functions which provide additive and subtractive properties to them [11]: CIE[2], L*ab* or L*uv* (Luminance and chrominance *uv* or *ab*), L*C*H* (Luminance, Chroma and Hue) or CIECAM02 (*CIE colour appearance model*)[12]. Other colour appearance models were created as a combination of others, i.e. HCL or L*C*H (hue, chroma and luminance)[13] inspired from HSL (descriptive/topological model) and CIE L*ab* (geometric model).

In the literature, there are different colour-naming approaches defined on different colour models: (i) a colour name descriptor was defined based on the CIE L*ab* colour model [14]; (ii) an approach for computational colour categorization and naming was formulated based on the CIE L*ab* colour model and fuzzy partitioning [15]; (iii) a computational approach for colour categorization and naming and extraction of colour composition was developed based on the CIE L*ab* and HSL colour models [16]; (iv) fuzzy colour categories were defined based on the Musell Colour Solid and the HCL colour model [17]; (v) an experimental study using the Munsell Colour Solid was carried out where the colour ranges reflecting the colour naming and perception of Turkish people for each colour term were obtained [18]; (vi) the dominant colour of a region (in HSV colour model) was converted into a set of 35 semantic colour names, some of them being related to natural scene images like *sky blue* or *grass green* [19]; (vii) an approach for colour-naming which introduced some semantic connotations, such as *warm/cold* or *light/dark* colours was defined on the HSL colour model [20]; (viii) twelve fundamental colours were defined on the CIE L*uv* colour space and semantic contrasts *warm/cold*, *light/dark* were given to them using Johannes Itten's theory of colour [21]; (ix) a computational approach for colour perception and colour-naming was defined based on the CIE XYZ and CIE L*ab* colour [22]; and

[2]CIE refers to the chromaticity diagram by the Commission Internationale de l'Eclairage

(x) a Colour Naming System (CNS) was formulated to quantize the HSL colour model into 627 distinct colours [23].

All the approaches described above provide evidence for the effectiveness of using different colour models and spaces for colour quantisation and naming. Note that names provided by the subjects are not affected by the specific way colours are encoded, and that quantisation algorithms can provide similar clusters based on similar data points. However, as Palmer [1] mentions: *The subjective experience of surface colour has a very different structure from that of physical light. All the surface colours experienced by a person with normal colour vision can be described in terms of just three dimensions: 'hue', 'saturation' and 'lightness'.* Thus, according to Palmer[1] and to Sarifuddin [13], the spatial distribution of colours in the HSL model is cognitive and intuitive for humans to divide it into intervals of values corresponding to colour names. Note also that HSL is broken down according to physiological criteria: hue refers to the pure spectrum colours and corresponds to the dominant colour as perceived by a human; saturation corresponds to the relative purity or the amount of white light that is mixed with hue; and luminance refers to the amount of light in a colour. Previous approaches also chose HSL colour model for their studies [23, 20, 16]. In contrast to them, the colour model based on HSL presented in this paper is designed to be generally adaptable and kept as simple and universal as possible since the most human beings can only manage a reduced number of colour names [3].

W3C[3] also mentions that additional advantages of HSL are that it is symmetrical to luminance and darkness which is not the case with HSV, for example. This means that: (i) HSV, when considering the value (V) at the maximum, it goes from saturated colour to *white* (which is not intuitive), whereas in HSL, the saturation (S) goes from fully saturated colours to *grey*; and (ii) in HSV, the value (V) only goes from *black* to the chosen hue, while in HSL, the luminance (L) always spans the entire range from black through the chosen hue to *white*. Therefore, the HSL colour space is suitable to be divided into intervals of values corresponding to colour names and also intuitive for adding semantic labels to these names in order to refer to the richness (saturation) or the brightness of the colour (luminance)[13].

Regarding similarity measures between colours, in the literature, different colour pixel similarity measures have been defined related to different colour models: (i) Euclidean distance is used in cubic representation colour models such as RGB or CIE L*ab* and occasionally in cylindrical colour models like L*C*H [13, 24]; (ii) a cylindrical distance was defined to obtain colour similarity on cylindrical and conical colour models like HSL, HSV and L*C*H [25]; (iii) similarity values based on the Fuzzy C-Means were defined to compare fuzzy colour categories based on the Musell Colour Solid in [17]; (iv) an interval distance was formulated for comparing colour names defined on HSL colour space [26]; and other formulae were defined for computing colour difference in L*C*H and CIECAM02 [27] and HCL [13]. As far as we are concerned, all the similarity measures presented above are calculated from the numerical values of the colour coordinates.

The approach presented in this paper obtains a similarity value between colour names, instead of between their exact colour coordinates, by taking into account the spatial relational structure of the colour model selected. To the best of our knowledge, in the literature, there are very few studies that try to calculate a similarity measure between colour names without using their pixel intensity values. Psychological studies based on surveys carried out on people [28, 29] have been the only attempts to obtain a similarity relation between colour names. In these studies,

[3] See the CSS3 specification from the W3C: `http://www.w3.org/TR/css3-colour/\#hsl-colour)`

participants were asked about *'which colour pair is the most similar: A and B or C and D?'* and diagrams of the psychological colour structure were built from the answers and then used to study colour symmetries and oppositions.

It is worth noting that the model proposed in this paper for colour naming and comparison besides taking into account cognitive perspectives and studies carried out previously in the literature, is also computational and it can be adapted to the requirements of any application.

3. The Computational QCD Model

The Computational QCD model translates the RGB colour channels into coordinates of the HSL [30] colour space (see Figure 1) in order to give a name to the colour displayed. From the HSL colour coordinates, a reference system for qualitative colour description is defined as follows:

$$QCRS = \{UH, US, UL, QC_{LAB1..M}, QC_{INT1..M}\}$$

where UH is the Unit of Hue; US is the Unit of Saturation; UL is the Unit of Luminance; and which holds: $0 \leq UH \leq 360$, $0 \leq UL \leq 50$ and $0 \leq US \leq 2 * UL$ (for the top cone) and $0 \leq UH \leq 360$, $50 \leq UL \leq 100$ and $0 \leq US \leq 200 - 2 * UL$ (for the bottom cone); and where $QC_{LAB1..M}$ refers to the qualitative labels related to colour distributed in M colour sets; and $QC_{INT1..M}$ refers to the intervals of Hue, Saturation and Luminance colour coordinates associated with each colour label of the M colour sets.

The HSL colour space distributes colours in the following way. The *rainbow colours* are located in the horizontal central circle. The colour luminance changes in the vertical direction,

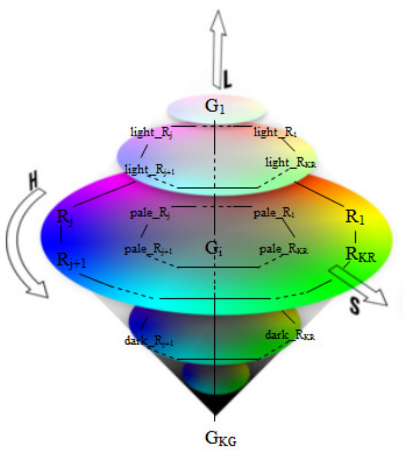


Figure 1: The QCD model of the HSL colour space

therefore *light rainbow colours* are located at the top, while *dark rainbow colours* are located at the bottom. The colour saturation changes from the boundary of the two cone bases to the axis of the cone bases and, therefore, *pale rainbow colours* are located inside the horizontal central circle. As a consequence of the changing colour saturation and luminance, the vertical axis locates the qualitative colours corresponding to the *grey scale*. According to this, the presented model for QCD considers[4] $M = 5$ colour sets: (1) grey colours, (2) rainbow colours, (3) pale rainbow colours, (4) light rainbow colours and, (5) dark rainbow colours, where the QC_{LAB_i} and QC_{INT_i}, for $i = 1, \cdots, 5$, are:

1. $\mathrm{QC}_{LAB_1} = \{\mathrm{G}_1, \mathrm{G}_2, \mathrm{G}_3, ..., \mathrm{G}_\ell\}$

   $\mathrm{QC}_{INT_1} = \{[0, \mathrm{g}_{ul_1}], (\mathrm{g}_{ul_1}, \mathrm{g}_{ul_2}], (\mathrm{g}_{ul_2}, \mathrm{g}_{ul_3}], ..., (\mathrm{g}_{ul_{\ell-1}}, 100] \in \mathrm{UL} \,/\, \forall\, \mathrm{UH} \in [0, 360] \wedge \forall\, \mathrm{US} \in [0, \min\{g_{us_{MAX}}, 2UL, 200-2UL\}]\}$,

   where ℓ colour names are defined for the grey scale in QC_{LAB_1} whose corresponding intervals of values in HSL are determined in QC_{INT_1}. All the colours in this set can take any value of hue, values of saturation between 0 and $\mathrm{g}_{us_{MAX}}$ and values of luminance (g_{ul_ℓ}) between 0 and 100, which determine the different colour names defined. Note that the saturation coordinate (US) determines if the colour corresponds to the grey scale or to the rainbow scale.

2. $\mathrm{QC}_{LAB_2} = \{\mathrm{R}_1, \mathrm{R}_2, \mathrm{R}_3, ..., \mathrm{R}_r\}$

   $\mathrm{QC}_{INT_2} = \{(\mathrm{r}_{uh_{r-1}}, 360] \wedge [0, \mathrm{r}_{uh_1}], (\mathrm{r}_{uh_1}, \mathrm{r}_{uh_2}], (\mathrm{r}_{uh_2}, \mathrm{r}_{uh_3}], ... ,(\mathrm{r}_{uh_{r-2}}, \mathrm{r}_{uh_{r-1}}] \in \mathrm{UH} \,/\, \forall\, \mathrm{UL} \in (\mathrm{r}_{ul_{MIN}}, \mathrm{r}_{ul_{MAX}}] \wedge \forall\, \mathrm{US} \in [r_{us_{MIN}}, \min\{100, 2UL, 200-2UL\}]\}$,

   where r colour names are defined for the rainbow scale in QC_{LAB_2} and are considered the more saturated ones. In QC_{INT_2}, their saturation can take values between $\mathrm{r}_{us_{MIN}}$ and 100, whereas their luminance can take values between $\mathrm{r}_{ul_{MIN}}$ and $\mathrm{r}_{ul_{MAX}}$. Thus, the different values of hue (r_{uh_r}) can take values between 0 and 360 and determine the colour names defined for this set.

3. $\mathrm{QC}_{LAB_3} = \{pale\_ + \mathrm{QC}_{LAB_2}\}$

   $\mathrm{QC}_{INT_3} = \{\, \forall\, \mathrm{UH} \in \mathrm{QC}_{INT_2} \,/\, \forall\, \mathrm{UL} \in (\mathrm{r}_{ul_{MIN}}, \mathrm{r}_{ul_{MAX}}] \wedge \forall\, \mathrm{US} \in (\mathrm{g}_{us_{MAX}}, \min\{r_{us_{MIN}}, 2UL, 200-2UL\}] \,\}$

   where r pale colour names are defined in QC_{LAB_3} by adding the prefix *pale_* to the colours defined for the rainbow scale, QC_{LAB_2}. The colour names defined in QC_{INT_3} have the same interval values of hue as rainbow colours (QC_{INT_2}). The lightness intervals also coincide, but they differ from rainbow colours in their saturation, which can take values between $\mathrm{g}_{us_{MAX}}$ and $\mathrm{r}_{us_{MIN}}$.

4, 5. $\mathrm{QC}_{LAB_4} = \{light\_ + \mathrm{QC}_{LAB_2}\}$

   $\mathrm{QC}_{INT_4} = \{\forall\, \mathrm{UH} \in \mathrm{QC}_{INT_2} \,/\, \forall\, \mathrm{UL} \in (\mathrm{r}_{ul_{MAX}}, 100] \wedge \forall\, \mathrm{US} \in [r_{us_{MIN}}, \min\{100, 2UL, 200-2UL\}] \,\}$

   $\mathrm{QC}_{LAB_5} = \{dark\_ + \mathrm{QC}_{LAB_2}\}$

   $\mathrm{QC}_{INT_5} = \{\forall\, \mathrm{UH} \in \mathrm{QC}_{INT_2} \,/\, \forall\, \mathrm{UL} \in (0, \mathrm{r}_{ul_{MIN}}] \wedge \forall\, \mathrm{US} \in [r_{us_{MIN}}, \min\{100, 2UL, 200-2UL\}] \,\}$

   where r light and dark colour names are defined in QC_{LAB_4} and QC_{LAB_5}, respectively, by adding the prefixes *dark_* and *light_* to the colour names in the rainbow scale (QC_{LAB_2}). The intervals of values for dark and light colour sets, QC_{INT_4} and QC_{INT_5}, respectively,

[4]Clearly, the QDC model can be broadly extended by choosing a major number of colour sets.

take the same values of hue as rainbow colours, QC_{INT_2}. The saturation intervals also coincide, but the luminance (UL) differs and determines the light or dark colours taking values between $r_{ul_{MAX}}$ and 100 or between r_{ul} and $r_{ul_{MIN}}$, respectively.

It is worth noting that the parameters ℓ (number of selected colour names for the grey scale) and r (number of chosen colour names for the rainbow scale) depend on the granularity that an expert needs in each scenario. The higher the values for these parameters, the more subjective the description, and the lower the values, the more universal the description.

As an example, taking as a reference the Natural Colour System (NCS) [31] the QCD model may establish three pairs of elementary colours (white-black, green-red and yellow-blue). According to that, the minimal values for these parameters would be assumed to be $\ell \geq 2$ (white and black) and $r \geq 4$ (green, red, yellow and blue). Therefore, the values $l = 2$ and $r = 4$ would be more universal than, for example, values of $l = 30$ where colour names such as *ivory* (a kind of white) could appear as needed in a more specific use case (i.e. snow expert or fashion designer).

According to Steels and Belpaeme [32], when grounding colour categories, multiple sources of constraints act: (i) constraints from embodiment, each visual sensory system can vary for every individual; (ii) constraints coming from the world, the individuals must be adapted to the environment and its statistical regularity has to be taken into account to reach viable performance; and (iii) constraints coming from cultural negotiation, or collective decisions made by population (i.e. a population may decide to combine blue and green categories, as many cultures have done). The QCD model can adapt its parameters ℓ and r to fulfill these constraints to the case of study.

4. Parameterising the QCD Model

In order to determine the interval of values associated to the Qualitative Colour Reference System, a test were carried out on 534 participants (students and teachers) at *Universitat Jaume I* and *Universidad de Sevilla* in Spain. A computer application was implemented which showed 10 different colours selected randomly and uniformly using their HSL coordinates. For each colour selected, participants were asked if they considered the colour to be in the *grey* or *rainbow* scale. For those colours classified in the *grey* scale, participants were asked if the colour was *white*, *light_grey*, *grey*, *dark_grey* or *black*, that is, $\ell = 5$. For those colours classified in the *rainbow* scale, participants were asked if the colour was *red*, *orange*, *yellow*, *green*, *turquoise*, *blue*, *purple* or *pink*, that is, $r = 8$, and if it was *light*, *pale* or *dark*. Thus, a total of 37 colour names were considered.

Let us justify the parameters selected: (i) $\ell = 5$ because the less saturated and extreme colours in luminance are *white* and *black* and, according to the M sets defined, there are two more gradations in lightness *light*- and *dark*- and one more in saturation *pale*-, which correspond to *light-grey*, *dark-grey*, and *grey*, respectively; and (ii) $r = 8$ since the rainbow/spectral colours are 7 and the majority of the participants of the test suggested to add also *pink*[5].

From the survey, a dataset with 5340 colour names and its corresponding HSL coordinates were obtained. Then, a supervised discretization algorithm, AMEVA [33], was used in order to calculate the classes of the intervals corresponding to each colour name. This algorithm was chosen because its main aim is to maximise the dependency relationship between the class labels,

[5]Note that the selected values for ℓ and r depend on the current use case and that different values of those parameters could have produced different outcomes in the survey.

Figure 2: Colour names in the grey scale and in the rainbow scale for the QCD model.

the colours, and the continuous values in HSL. In other words, the AMEVA algorithm obtains the intervals of values that best fit the colour names provided by the judgments of the participants measured from the contingency coefficient between colours and intervals. Note also that the AMEVA algorithm discretises each variable independently from the others. However, the dependency constraint of the unit of Saturation and the unit of Lightness in the HSL colour space has been also taken into account.

As a result, Table 1 shows the values extracted by AMEVA for parameterising the QCD model, taking into account the topological structure of the HSL colour space showed by the QCRS, and Figure 2 shows the colour values assigned to each colour name, which correspond to the central value of each interval in HSL.

Figure 3 shows that the QCD model gives the same colour category to different colour intensities in the same way as suggested by participants. It is straightforward to see that most of the people may agree to name any of the colours in each grid with the name given by the QCD model.

5. Analysing the Relational Structure of the QCD Model

The relational structure of the QCD model is studied by analysing the conceptual neighbourhood of the qualitative colours defined. Freksa [34] defined that two qualitative concepts in space are conceptual neighbours if *'one can be directly transformed into another by continuous deformation'*. This definition is applied to the colour space HSL. Let us exemplify this, the colours

Table 1: HSL intervals for colour names.

	Colour Name	**UH**	**US**	**UL**
	black			[0, 20)
	dark_grey			[20, 30)
QC_{LAB_1}	*grey*	[0, 360]	$[0, \min\{20, 2UL, 200-2UL\}]$	[30, 40)
	light_grey			[40, 80)
	white			[80, 100]
	red	(335, 360] ∧ [0, 20]		
	orange	(20, 50]		
	yellow	(50, 80]		
QC_{LAB_2}	*green*	(80, 160]	$(50, \min\{100, 2UL, 200-2UL\}]$	(40, 55]
	turquoise	(160, 200]		
	blue	(200, 239]		
	purple	(239, 297]		
	pink	(297, 335]		
	pale_red	(335, 360] ∧ [0, 20]		
	pale_orange	(20, 50]		
	pale_yellow	(50, 80]		
QC_{LAB_3}	*pale_green*	(80, 160]	$(20, \min\{50, 2UL, 200-2UL\}]$	(40, 100]
	pale_turquoise	(160, 200]		
	pale_blue	(200, 239]		
	pale_purple	(239, 297]		
	pale_pink	(297, 335]		
	light_red	(335, 360] ∧ [0, 20]		
	light_orange	(20, 50]		
	light_yellow	(50, 80]		
QC_{LAB_4}	*light_green*	(80, 160]	$(50, \min\{100, 2UL, 200-2UL\}]$	(55, 100]
	light_turquoise	(160, 200]		
	light_blue	(200, 239]		
	light_purple	(239, 297]		
	light_pink	(297, 335]		
	dark_red	(335, 360] ∧ [0, 20]		
	dark_orange	(20, 50]		
	dark_yellow	(50, 80]		
QC_{LAB_5}	*dark_green*	(80, 160]	$(20, \min\{100, 2UL, 200-2UL\}]$	(0, 40]
	dark_turquoise	(160, 200]		
	dark_blue	(200, 239]		
	dark_purple	(239, 297]		
	dark_pink	(297, 335]		

red and *orange* are conceptual neighbours since a continuous change in hue causes a direct transition from *red* to *orange*. However, the colours *yellow* and *red* are not conceptual neighbours because a continuous transformation of hue from *red* gets the colour *orange* in between. Other

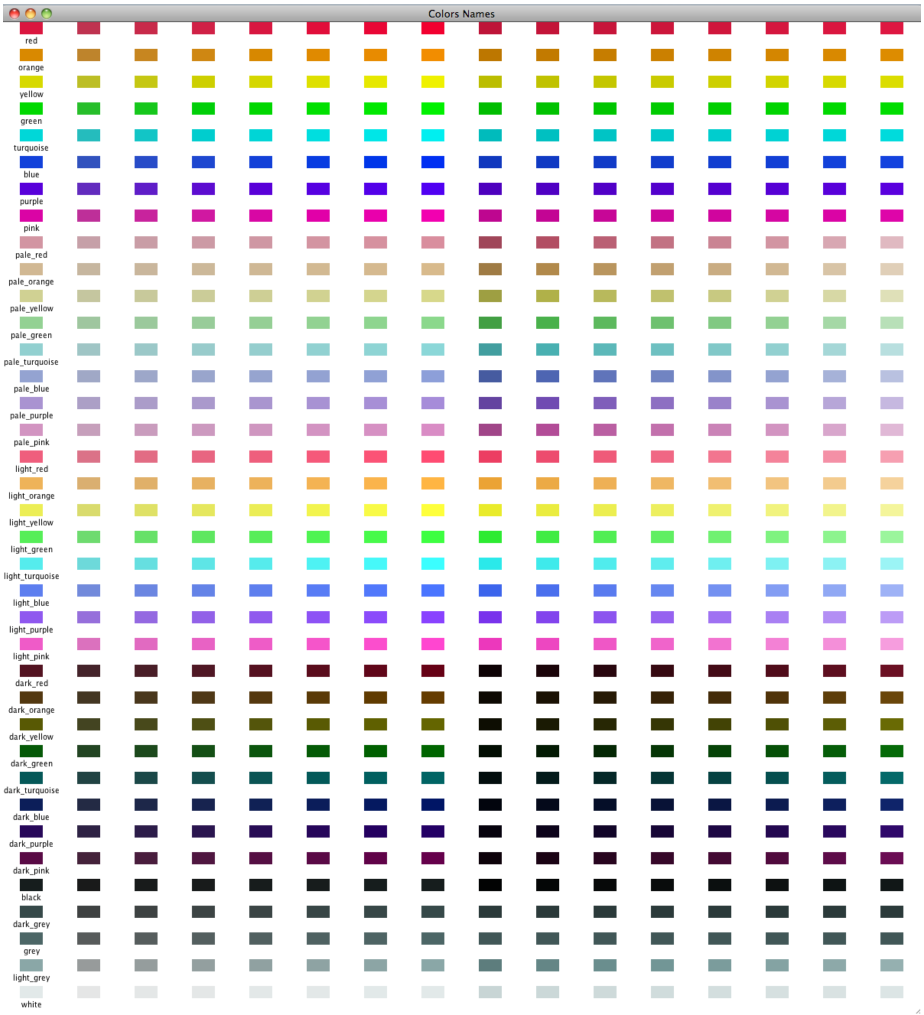


Figure 3: Different HSL values corresponding to the same colour name for the QCD model

conceptual neighbours of *red* which are derived from continuous transformation in lightness are *dark-red* and *light-red* and the conceptual neighbour of *red* varying the saturation is *pale-red*.

Therefore, a conceptual neighbourhood diagram (CND) can be derived which contains: (i) nodes that map to a set of individual relations defined on intervals; and (ii) paths connecting pairs of adjacent nodes that map to continuous transformations which can have weights assigned in order to establish priorities. According to this, a CND for the computational QCD model has been built and it is shown in Figure 4. The nodes of this CND correspond to the colour names, whereas the path connecting neighbouring colours are drawn by lines which are assigned weights to establish priorities.

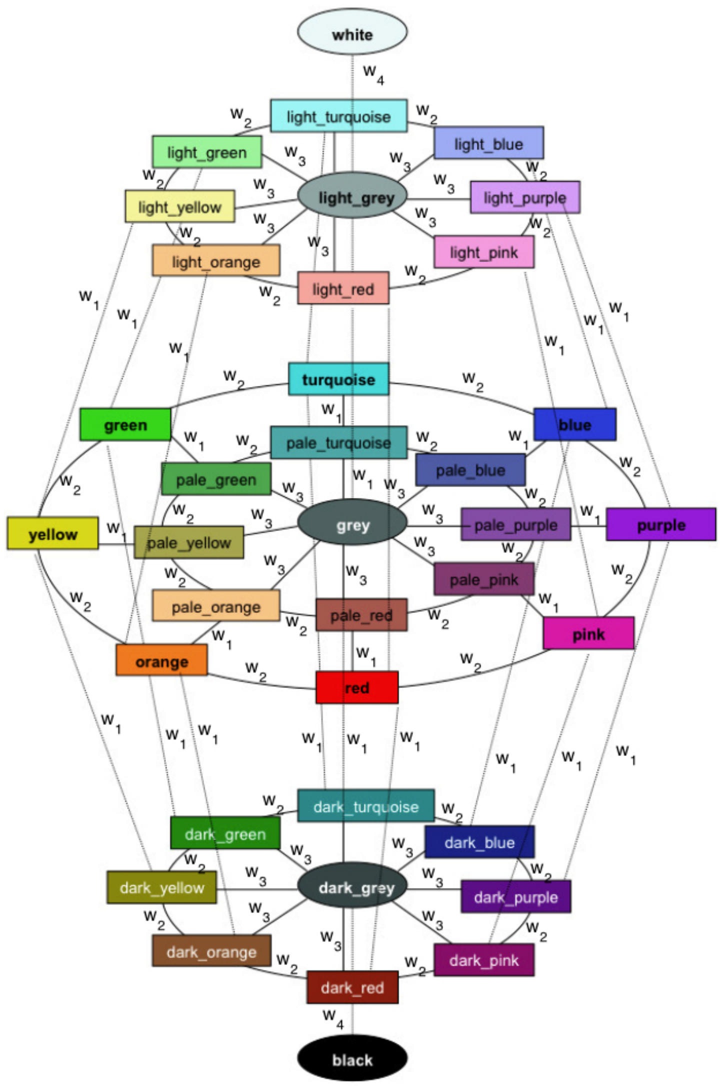


Figure 4: A Conceptual Neighbourhood Diagram for the QCD model. Note that the colour names are located where the centroid of the HSL colour intervals is and that w_i denotes the importance of the transitions/changes in HSL.

6. A Similarity Measure for the QCD Model

The dissimilarity between qualitative colours in the QCD model, denoted by $dsColour(\cdot,\cdot)$, is calculated as the minimal path between the nodes of the CND in Figure 4. In this CND, the paths connecting pairs of adjacent nodes that map to continuous transformations can be assigned the following positive weights in order to establish priorities:

- w_1 is the weight assigned to the transition between a colour name and the same colour name with a semantic prefix (*pale_, light_, dark_*), that is, to transitions that do not involve changes in the hue colour coordinate. For example: *dsColour(red, light_red) = dsColour(grey, dark_grey)* = w_1.
- w_2 is the weight assigned to the transitions between colour names in the rainbow scale with or without a semantic prefix (*pale_, light_, dark_*). For example: *dsColour(pink, red) = dsColour(pale_pink, pale_red)* = w_2.
- w_3 is the weight assigned to the transition between the colours in the grey scale and the *light*, *pale* and *dark* colours on the rainbow scale. For example: *dsColour(pale_red, grey) = dsColour(light_yellow, light_grey) = dsColour(dark_blue, dark_grey)* = w_3.
- w_4 is the weight assigned to the transitions between *black* and *white* colour names and the colours in the grey scale. For example: *dsColour(black, dark_grey) = dsColour(white, light_grey)* = w_4.

According to the importance of these transitions, the following relations are hold:

- w_1 is given to the changing transition between a colour name and the same colour name (same hue) but different lightness or saturation, whereas the w_2 is given to the changing transitions between different colour names (different hues). From a cognitive point of view, the difference in colour perception is higher when the hue changes that when it does not; in fact, not perceiving the difference between some hues is considered a disease (i.e. colourblindness). Hence $w_1 \leq w_2$ is considered.
- w_3 is given to the changing transition between a colour name (denoted by any hue) and another colour name denoting the absence of hue (grey scale). From a cognitive point of view, the difference between perceiving colours (i.e. normal vision) to not perceiving any of them (i.e. acromatopsia) [35] is more significant than the difference between normal vision and confusing slightly different hues (i.e. red-green colourblindness). Hence $w_2 \leq w_3$ is considered.
- w_4 is given to the changing transition between *white* (full light)/*black* (absence of light) and another colour name in a grey scale. From a cognitive point of view, the change of having only two distinctions in light is more significant than having a range of grey perception; hence $w_3 \leq w_4$ is considered.

Therefore, the priorities established must verify: $0 < w_1 \leq w_2 \leq w_3 \leq w_4$.

Hence, given two qualitative colours, denoted by QC_A and QC_B, a similarity between them, denoted by $SimQCD(QC_A, QC_B)$, is defined as:

$$SimQCD(QC_A, QC_B) = 1 - \frac{dsColour(QC_A, QC_B)}{MaxDsColour} \tag{1}$$

where $dsColour(QC_A, QC_B)$ denotes the previously defined dissimilarity and $MaxDsColour$ denotes the maximum dissimilarity for all colour names.

The main properties of this similarity measure are:

- Symmetry: $SimQCD(QC_A, QC_B) = SimQCD(QC_B, QC_A)$
- Upper and lower bounds: $0 \leq SimQCD(QC_A, QC_B) \leq 1$
- Intuitive: $SimQCD(QC_A, QC_B) = 0$ means that $dsColour(QC_A, QC_B) = MaxDsColour$, that is, both colours are as different as possible. $SimQCD(QC_A, QC_B) = 1$ means that $dsColour(QC_A, QC_B) = 0$, that is, both colours are the same.

Given some qualitative colours, the model can also calculate relative colour comparisons, such as:

- *'Is QC_A darker/lighter than QC_B?'* by calculating and proving whether:
  $SimQCD(QC_A, black/white) > SimQCD(QC_B, black/white)$
- *'Is QC_A bluer/redder/etc. than QC_B?'* by calculating and proving whether: $SimQCD(QC_A, rc) > SimQCD(QC_B, rc)$, where rc = {*blue*/*red*/etc.}

7. Parameterising the SimQCD Model

The *SimQCD* calculus is parameterised by assigning, as a baseline, the following values to weights: $w_1 = 1$, $w_2 = 3$, $w_3 = 5$ and $w_4 = 6$. Hence, $MaxDsColour = 14$ which is given between *black* and *white* colours.

The adequacy of this parameterisation is tested by:

- comparing the different HSL coordinates which are assigned the same colour name (Figure 3); and
- calculating all the similarity values obtained between all the qualitative colours defined with the aim of testing arrangements of the most similar colours.

Some results for the 37 representative colour names are given in Figure 5 and Figure 6. In these figures, the representative colour name is given first; and then, the 10 most similar colours are arranged according to *SimQCD* showing: the representative colour display, the colour name and the similarity value obtained.

From the gradation of colours built according to the similarity values obtained by *SimQCD* have some intuitive properties are extracted:

- the null similarity is given between *white* and *black*.
- the similarity given between any *rc* and *black/white* or any *pale rc* and *black/white* is the same.
- the same similarity is given between any *light rc* and *white* and any *dark rc* and *black*.
- the same similarity is given between any *light rc* and *dark* and any *light rc* and *black*.
- the similarity given between any *rc* and the same *dark*, *pale* or *light rc* is the same.
- the same similarity is given between any *prefix* (*pale*, *dark* or *light*) of the same *rc*.
- the similarity given between any *pale rc* and *grey*, and between any *light rc* and *light_grey*, and between any *dark rc* and *dark_grey* is the same.
- any *light rc* is more similar to *white* than any *pale rc* to *white* and, in the same way, any *dark rc* is more similar to *black* than any *pale rc* to *black*.

Figure 5: Similarity calculus applied to compare 18 qualitative colours defined in the QCD model. The ten most similar colours are displayed.

8. The Complementary Colour in the QCD Model

Complementary colours are pairs of colours that are of *opposite* hue in a colour model and were defined first by Goethe in his *Theory of Colours* [36]. The exact hue *complementary* to a given hue depends on the colour model applied.

In colour theory, two colours are called complementary if, when mixed in the proper proportion, they produce a neutral colour (*grey*, *white*, or *black*). In roughly-perceptual colour models,

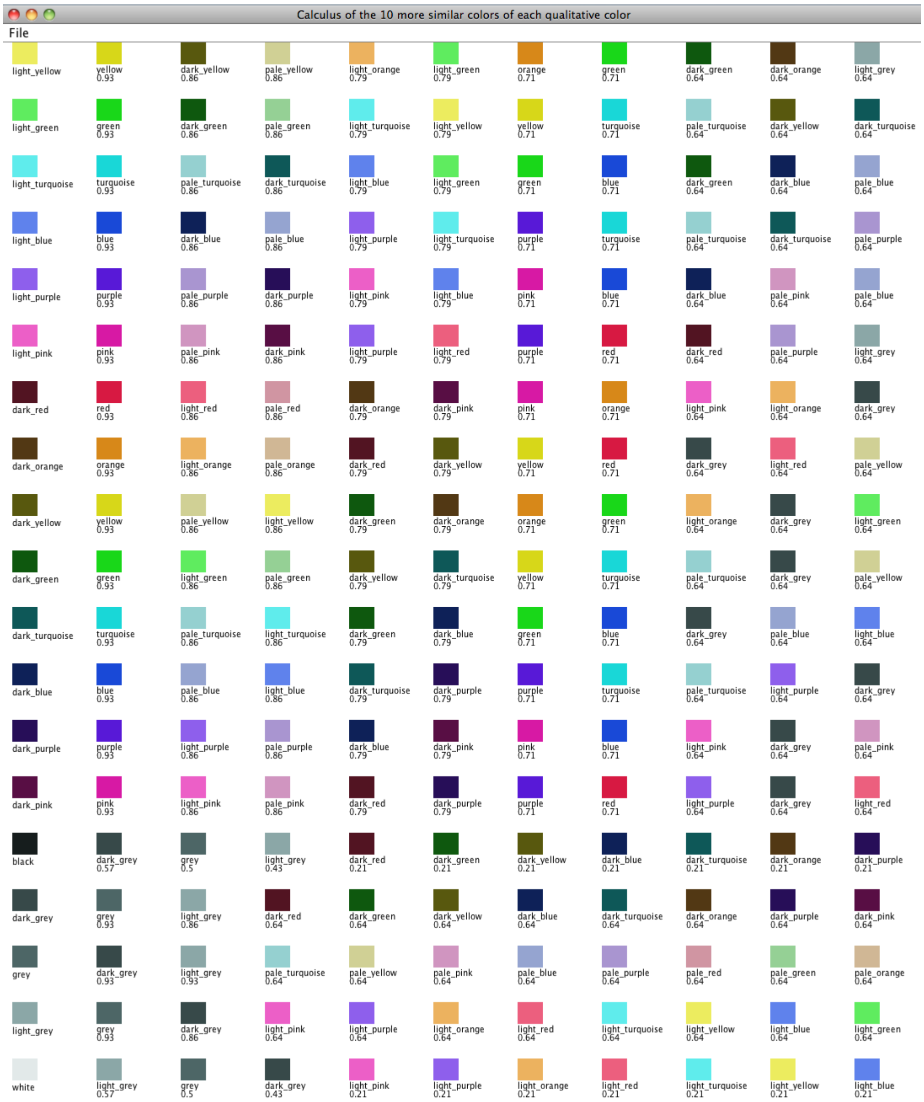


Figure 6: Similarity calculus applied to compare the rest of the 19 qualitative colours defined in the QCD model. The ten most similar colours are displayed.

the neutral colours lie along a central axis, as in HSL colour space.

For the colours in the rainbow scale in the QCD model, the addition of two complementary colours produces the colour *white*. The colour coordinates selected for calculating the complementary of those colours were those corresponding to the centre of each wedge since Berlin and Kay [41] demonstrated that humans determined prototypical colours as the centre of colour categories. In HSL colour space, the colour *white* is determined by the coordinates $(uh, us, 100)_{HSL} / uh \in [0, 360], us \in [0, 100]$.

Hence, given a qualitative colour defined by the centre of its wedge (centroid) in the QCD model, $QC_A = (H,S,L)_{HSL}$, the complementary colour is calculated as:

$$\overline{QC_A} = ((180+H)\%360, S, 100-L)$$

The calculus of the complementary colours in the QCD model has been tested and the results are shown in Figure 7. The complementary colour verifies two important properties:

- $\overline{\overline{QC_A}} = QC_A$; and
- $SimQCD(\overline{QC_A}, QC_A)$ is the same as the colour with the lowest similarity inside the same colour scale (QC_{INT_i}).

Figure 7: Complementary colours in the QCD model and the *SimQCD* measure between them.

9. Similarity of Compositions involving Different Qualitative Colours

The similarity measure defined between the qualitative colours in the QCD model is used to compute the similarity of two compositions (digital images) based on the colours appearing in them and their percentage of appearance.

Let us denote the set of the 37 representative colour names of the QCD model as: $\mathscr{C} = \{QC_1, \cdots, QC_{37}\}$. Thus, the similarity $SimQCD : \mathscr{C} \times \mathscr{C} \longrightarrow [0,1]$ provides a matrix

$$S = \left\{SimQCD(QC_i, QC_j)\right\}_{i,j=1}^{37}$$

which is symmetric and whose main diagonal contains 1 values.

Let us consider $\mathscr{Y}$ as the set of the colour compositions/images to compare. If *Image* represents a colour composition, the system obtains a colour histogram:

$$Image = (f_1, f_2, \cdots, f_{37})$$

where f_i corresponds to the percentage of the colour QC_i within the *Image* ($f_i \geq 0$). Therefore, each image is assigned a unique vector,

$$\mathscr{Y} \longrightarrow \mathbb{R}^{37}$$

that is, $Image \equiv \mathbf{I}$ where $\mathbf{I} \in \mathbb{R}^{37}$. Note that two images or colour compositions are equal in the system presented if they have the same representation as $\mathbb{R}^N$ vector.

In order to define a similarity measure, let us consider the following matrix S^* associated to S and defined as follows:

$$S^* = \{s^*_{ij}\}_{i,j=1}^{37}$$

where[6]

$$s^*_{ij} = \begin{cases} 0.5 \cdot SimQCD(QC_i, QC_j) & i \neq j \\ SimQCD(QC_i, QC_j) & \text{otherwise} \end{cases}$$

Thus, a Quadratic Form[7] is considered as follows:

$$QF : \mathbb{R}^{37} \longrightarrow \mathbb{R}, \qquad QF(\mathbf{x}) = \mathbf{x}\, S^*\, \mathbf{x}^t$$

and given an image $Image = (f_1, f_2, \cdots, f_{37})$ is obtained that

$$QF(Image) = \sum_{i=1}^{37} \sum_{j=1}^{37} f_i\, f_j\, s^*_{ij}$$

The S^* matrix is defined positive since all its eigenvalues are positive (see Table 2). Therefore,

Table 2: Eigenvalues of the S^* matrix

Eigenvalues	Number	Eigenvalues	Number
0.4793	1	0.5000	25
0.5154	1	0.6583	1
0.8456	1	1.0021	2
1.0775	1	1.2756	1
1.2973	1	3.4265	2
9.4940	1	Total	37

QF defines a norm in $\mathbb{R}^{37}$ as follows: $\|\mathbf{x}\| = \sqrt{QF(\mathbf{x})}$ for any $\mathbf{x} \in \mathbb{R}^{37}$, and hence, a 'quasi'-distance[8] in $\mathscr{Y}$ is defined as:

$$d : \mathscr{Y} \times \mathscr{Y} \longrightarrow \mathbb{R}$$

$$d(Image1, Image2) = \|\mathbf{I}_1 - \mathbf{I}_2\|$$

where $Image1 = \mathbf{I}_1 = (f_1, \cdots, f_{37})$ and $Image2 = \mathbf{I}_2 = (f'_1, \cdots, f'_{37})$.

Furthermore, it holds that

$$0 \leq \|\mathbf{I}_1 - \mathbf{I}_2\|^2 = \|\mathbf{I}_1\|^2 + \|\mathbf{I}_2\|^2 - 2 \sum_{i,j=1}^{37} f_i f'_j s^*_{ij} \leq 1 + 1 = 2$$

since s^*_{ij}, $f_i\, f_j \geq 0$ for any i, j, and:

$$\begin{aligned} \|\mathbf{I}\|^2 &= \textstyle\sum_{i=1}^{37} \sum_{j=1}^{37} f_i\, f_j\, s^*_{ij} \leq \sum_i^{37} \sum_{j=1}^{37} f_i\, f_j \\ &= \left(\textstyle\sum_{i=1}^{37} f_i\right) \left(\textstyle\sum_{j=1}^{37} f_j\right) = 1 \end{aligned}$$

[6] A 0.5 factor is needed in order to avoid the duplicity of $f_i \cdot f_j$ when $i \neq j$.

[7] $\mathbf{x}^t$ means the transpose vector of $\mathbf{x}$.

[8] The distance condition $d(\mathbf{x}, \mathbf{y}) = 0 \Rightarrow \mathbf{x} = \mathbf{y}$ is not true.

From the distance, $d(\cdot,\cdot)$, a similarity measure between two images regarding only their colour compositions $\mathbf{I}_1$ and $\mathbf{I}_2$ is obtained as follows:

$$SimQCDI(\mathbf{I}_1,\mathbf{I}_2) = 1 - \frac{d(\mathbf{I}_1,\mathbf{I}_2)}{\sqrt{2}}$$

The main properties of the *SimQCDI* similarity are:

- $0 \leq SimQCDI(\mathbf{I}_1,\mathbf{I}_2) \leq 1$
- If $\mathbf{I}_1 = \mathbf{I}_2$ then $d(\mathbf{I}_1,\mathbf{I}_2) = 0$ and, hence $SimQCDI(\mathbf{I}_1,\mathbf{I}_2) = 1$, that it, the maximum similarity.
- $SimQCDI(\mathbf{I}_1,\mathbf{I}_2) = SimQCDI(\mathbf{I}_2,\mathbf{I}_1)$, that is, the similarity is symmetric.

10. Experimentation

Experiments have been carried out to evaluate the model for colour naming (QCD) and the similarity measures defined (*SimQCD* and *SimQCDI*) using art compositions as the scenario (Section 10.1). Moreover, a survey which included images from the scenario was carried out (Section 10.2) and the similarity results obtained after comparing all the images in the testing scenario using the *SimQCDI* (Section 10.3) where compared to the results obtained by the survey (Section 10.4).

10.1. Scenario: Art Compositions

The scenario proposed for the experimentation consists on comparing art compositions taking into account only the colours in the paintings. The following painters were selected because of their different countries of origin, techniques and periods:

- Doménikos Theotokópoulos (1541-1614), *el Greco* as he was usually nicknamed, was a Greek painter in the Spanish Renaissance.
- Diego Velázquez (1599-1660) was one of the most important painters of the Spanish Golden Age in the contemporary Baroque period.
- Joan Miró (1893-1983) was a Catalan-Spanish painter, sculptor, and ceramicist who earned international acclaim and whose work was interpreted as Surrealism.
- Salvador Dalí (1904-1989) was a prominent Catalan-Spanish surrealist painter.
- Friedensreich Hundertwasser (1928-2000) was an Austrian artist who created the Transautomatism, a kind of surrealism, focused on the viewer's fantasy rather than an objective interpretation.

Five paintings of each author from the following digital on-line galleries: Wikipedia[9], Museo del Prado[10], Museo Reina Sofía[11], Museo Frieder Burda[12], Fundació Joan Miró[13], Fundació Salvador Dalí[14], Hundertwasser Foundation[15], and Kunst für alle[16] are shown in Figure 8.

The main aim of the experimentation is to determine the colour similarity in art compositions even when the painters use different techniques/periods (i.e. surrealism vs. baroque). And also to identify differences in colour inside the same techniques/periods. The *SimQCDI* can analyse two compositions based on: (i) all the different qualitative colours within the images; and (ii) the percentage of appearance in them.

10.2. Cognitive Test: Survey on Art Compositions

A user test was developed in order to research on the following hyphoteses:

- H1 If the QCD developed is cognitive when comparing paintings of the same painter, that is, if the similarity given between the paintings belonging to a given painter can grade the paintings in the same order as participants in the survey did;
- H2 If the QCD may be used to distinguish perceptually between paintings of different painters, that is, if the similarity provided is high when participants think that 2 paintings are similar and not otherwise;
- H3 If the QCD can manage visual effects as the background colour of the paintings, or colour contrasts as participants in the survey did.

This survey[17] was spread out as a Google Docs form inside a Google Sites and collected 109 responses. Approximately the 60% (65/109) of the participants were male and the 40% (44/109) were female and their ages were between 26 and 35 years old. Most of the participants spent between 9 and 12 minutes to answer the test. Around the 48% (52/109) of the participants considered their level of expertise in colour discrimination as very low, 41% (52/109) as medium and only 11% (12/109) as high. Most of the participants had a degree, master degree or PhD.

An example of a question in the survey is the following: *"Which two images in Fig. 8 are more similar considering only the colour distribution: (a) D4 and D1; (b) D4 and D5; or (c) D1 and D5?"* And results obtained were: 37% of the participants thought that the most similar art compositions in terms of colour are A and C; 40% of the participants voted for B and C, and 23% of the participants chose A and B. This example shows that participants did not see clearly any remarkable difference between any pair of these art compositions, maybe because all these compositions are by the same author, Dalí.

It is easier for computational approaches to be objective or not influenced by the shapes and context in the compositions. For this reason, the similarities between the compositions in Figure 8 are computed in the next section and then compared to the results of this survey in the discussion.

[9] `http://www.wikipedia.org`
[10] `http://www.museodelprado.es/coleccion/galeria-on-line/`
[11] `http://www.museoreinasofia.es`
[12] `http://www.museum-frieder-burda.de`
[13] `http://www.fundaciomiro-bcn.org`
[14] `http://www.salvador-dali.org`
[15] `http://www.hundertwasser.at/`
[16] `http://www.kunst-fuer-alle.de/`
[17] `https://sites.google.com/a/uji.es/colour-image-simi/`

Table 3: Similarity values between the art compositions in Figure 8.

	D2	D3	D4	D5	G1	G2	G3	G4	G5	H1	H2	H3	H4	H5	M1	M2	M3	M4	M5	V1	V2	V3	V4	V5
D1	77.09	71.77	76.83	77.91	75.77	70.46	69.23	78.87	79.51	79.68	76.11	73.87	78.76	74.30	71.36	65.29	61.76	79.34	55.25	60.13	64.50	74.30	69.25	68.33
D2		69.68	90.84	82.42	78.88	79.09	81.73	89.08	79.78	74.21	74.59	77.97	79.00	76.51	62.05	67.00	59.83	85.73	44.89	65.34	71.42	87.83	80.16	84.91
D3			70.36	66.66	65.40	69.82	64.44	68.58	80.31	72.40	71.79	66.73	72.90	68.22	61.01	56.80	53.07	67.29	49.23	79.93	79.25	68.22	66.30	65.70
D4				79.50	76.89	80.27	81.60	84.97	77.47	72.98	74.66	76.35	79.20	76.35	63.16	64.12	58.21	81.39	43.06	66.42	72.16	84.90	81.70	82.90
D5					75.93	73.90	69.26	80.04	79.16	69.76	68.50	69.37	71.14	70.06	60.07	68.37	60.43	89.43	47.27	57.13	60.87	76.80	72.47	70.56
G1						84.70	80.93	84.89	79.20	74.25	72.46	77.35	83.05	81.43	56.80	71.97	63.01	80.42	51.71	54.91	60.98	81.25	84.43	77.71
G2							80.55	79.27	79.79	71.38	73.14	72.70	81.52	79.99	55.54	65.88	59.08	75.86	46.14	61.83	67.98	80.51	93.56	79.71
G3								85.40	73.29	73.00	71.91	82.09	83.75	81.79	55.85	65.67	57.45	74.42	44.26	60.76	67.61	84.29	83.60	86.18
G4									80.99	77.62	76.06	83.59	82.80	80.97	61.71	71.49	62.58	86.72	50.45	61.71	67.45	88.61	80.16	83.74
G5										81.41	80.92	74.77	80.98	79.73	62.66	68.27	61.78	81.66	55.89	67.36	72.62	79.77	76.36	74.45
H1											89.28	79.80	82.86	84.02	64.01	63.39	60.45	72.23	53.19	63.41	70.49	74.38	70.19	70.94
H2												76.80	82.90	84.32	63.06	61.09	57.97	71.71	48.62	64.88	73.53	74.82	72.21	72.45
H3													81.95	84.52	61.23	66.59	60.07	74.79	49.85	60.54	66.24	79.24	73.36	77.21
H4														87.87	63.76	66.12	60.09	75.19	51.17	64.00	70.83	80.05	80.86	78.58
H5															59.44	66.30	60.62	74.33	50.50	60.00	67.63	78.57	79.61	76.80
M1																52.36	51.22	60.72	46.21	52.50	55.56	58.88	54.67	55.20
M2																	83.63	71.84	50.07	46.69	50.90	68.81	64.88	64.33
M3																		62.52	45.85	43.96	48.19	60.65	58.18	56.65
M4																			49.48	58.67	63.16	83.92	74.70	76.66
M5																				32.12	36.04	47.62	44.90	41.94
V1																					87.52	63.10	59.75	63.98
V2																						70.04	66.64	71.90
V3																							81.37	89.63
V4																								81.58

10.3. Computational Test: Similarity between Art Compositions

The colours in all the paintings in Figure 8 are extracted and interpreted qualitatively according to the QCD model. Then, the *SimQCDI* similarity measure was computed for comparing: (i) pictures by the same artist, to try to find colour similarities between them; and (ii) pictures by different artists, to analyze if a similarity only based on colour may be used to differentiate between artists.

Results obtained when comparing the art compositions in Figure 8 are given in Table 3. The mean and the standard deviation of the similarities are given in Table 4.

Table 4: The mean and standard deviation of the similarities among art compositions by authors.

	Dalí	*Greco*	*Hundertwasser*	*Miró*	*Velázquez*
Dalí	76.31 ± 6.75	76.42 ± 5.99	73.54 ± 3.72	63.02 ± 11.62	72.46 ± 08.32
Greco	—	80.90 ± 3.40	78.43 ± 4.03	63.59 ± 10.61	75.10 ± 10.02
Hundertwasser	—	—	83.43 ± 3.41	62.27 ± 07.65	72.03 ± 06.11
Miró	—	—	—	57.39 ± 11.72	55.59 ± 12.09
Velázquez	—	—	—	—	73.55 ± 10.17

Regarding the comparison of art pieces by the same author in terms of the art compositions selected, the following statements can be extracted:

- The artist who more often repeats the same palette of colours in similar proportions is Hundertwasser since the similarity obtained between them is 83.43% with a low variability ($\pm$3.41). Note that, very similar *red*, *yellow*, *blue* and *green* and *dark* colours appear in almost all the compositions selected. This also happens for the selected pictures by Greco, which obtain a similarity value of 80.90% with a low variability ($\pm$3.40) between them.

- The artist who uses a large palette of colours here is Miró since the similarity obtained between them is 57.39% with a high variability ($\pm$11.72). Note that the art compositions selected have different background colours, which may affect colour similarity.

- The selected art compositions by Dalí obtain a similarity measure of 76.31% which is quite similar to those by Velázquez 73.55%. This indicates that the colours used by these authors in their art compositions are similar and that they also use them in similar proportions. However, they obtain different variability of colours in their compositions. The variability obtained by Dalí is higher (± 6.75) than the one obtained by Hundertwasser and Greco. Note that Dalí usually uses *blue* and *yellow* colours contrasting with *greys* of different lightness. The variability obtained by Velázquez is higher (± 10.17) who also uses *dark* colours contrasting with *blue* and *yellow* but also *red* colours.

With respect to the comparison of art pieces by different authors, it is shown that:

- The *red*, *yellow*, *blue* and *green* colours contrast with *dark* colours in art compositions by both authors, Greco and Hundertwasser, and this produces quite high similarity values between their art pieces (78.43 ± 4.03).

- Hundertwasser obtains higher similarity values when comparing his own art compositions among them (83.43%) than when comparing his art compositions to those produced by other authors (73.54%, 78.43%, 62.27% and 72.03%). The same fact is obtained by Greco's selected art compositions: 80.90% versus 76.42%, 78.43%, 63.59% and 75.10%.

- It worth noting that Miró obtains lower similarity values when comparing his own art compositions (57.39%) than when comparing those with art compositions by Dalí (63.02%), Greco (63.59%) and Hundertwasser (62.27%).

- In fact, the painters with less similar art compositions are Miró and Velázquez (55.59%), and Miró taking into account their own paintings (57.39%).

With respect to the comparison of specific art pieces across different authors, it is shown that:

- The composition M4 by Miró obtains high similarity values to some art pieces by Dalí (85.73% - 89.43%), because of its *grey* background (see Section 11.1). It also obtains a high similarity to V3 by Velázquez (83.92%) and G4 by Greco (81.66%) because of the similar amount of *blue* and *grey* colours in both compositions.

- The compositions V3-V5 by Velázquez obtain high similarities to the art pieces by Greco: the appearance of *blue*, *red*, *yellow*, *grey* and *dark* colours is common in most of the compositions.

- The most similar pictures are G2 by Greco and V4 by Velázquez since a 93.56% of similarity is attained. On the other hand, the least similar pictures are M5 by Miró and V1 by Velázquez since a 32.12% of similarity is attained.

The descriptions above imply that, considering two art compositions, only using the *SimQCDI*, it cannot be determined if they were painted by the same artist or not. This could be achieved by studying the authors' palette and formulating a classification algorithm which make use of learning techniques such as support vector machines (SVMs) [37], neural networks [38], tree decisions algorithms i.e. C4.5 [39], and so on.

10.4. Comparing the Similarity Results to the Survey Results

The results obtained by the computational models QCD and *SimQCDI* are compared with the main results provided by the participants of the survey. To simplify, the results obtained in the survey are presented in each corresponding item where they are discussed.

The survey asked the participants which pair of art pieces by the same authors were more similar according to their colours:

- When comparing the art pieces D1-D4-D5, the results in Table 5 were obtained. From these data, the ideal results would be to obtain the couples (76.83, 23), (77.91, 37) and (79.50, 40), that is, the higher the similarity, the higher amount of votes. However, this result is not obtained. The three art pieces are very similar in colours and the participants are choosing their favorite pairs following personal criteria. Nevertheless it can be concluded

Table 5: Survey results and *SimQCDI* values for D1-D4-D5.

	SimQCDI		% of votes	
	D4	D5	D4	D5
D1	76.83	77.91	23	40
D4	–	79.50	–	37

that the similarities provided by *SimQCDI* are near to the opinion of the most participants.

- When comparing the art pieces G1-G2-G3, the answers gathered were those in Table 6. In this case, the *SimQCDI* similarity agrees completely with the participants of the survey, since the difference in similarity between (80.93, 16) and (80.55, 17) is not very significant.

- When comparing the art pieces H1-H2-H4, the votes were those indicated in Table 7. In this case, all the similarities obtained by *SimQCDI* are very high, and they agree with the opinion of the participants of the survey: the higher the similarity in colours between art pieces, the higher number of votes.

The survey also asked the participants to compare pairs of art pieces by different authors and the following results were provided:

- When comparing the art pieces V1-G2 versus V1-D4, the results in Table 8 were obtained. The 50% of the participants chose each pair equally, which coincides with the similarity values obtained, which are relatively close.

Table 6: Survey results and *SimQCDI* values for G1-G2-G3.

	SimQCDI		% of votes	
	G2	G3	G2	G3
G1	84.70	80.93	67	16
G2	–	80.55	–	17

Table 7: Survey results and *SimQCDI* values for H1-H2-H4.

	SimQCDI		% of votes	
	H2	H4	H2	H4
H1	89.28	82.86	46	22
H2	–	82.90	–	32

Table 8: Survey results and *SimQCDI* values for comparing V1-G2 and V1-D4 pairs.

	SimQCDI	% of votes
V1-G2	61.83	50
V1-D4	66.42	50

- When comparing the art pieces D1-M2 versus D1-H2, the results in Table 9 were obtained. In this case, note that an inverse control-question was made, that is, which pair of art pieces was less similar. The opinion of the participants agrees with the dissimilarity values calculated as $1 - SimQCDI$. The fact that the participants noticed when the survey was asking 'more' or 'less' similar pairs confirms that they did the survey thoughtfully. Therefore, according to these answers, the survey results were validated.

Table 9: Survey results and *SimQCDI* values for comparing D1-M2 versus D1-H2.

	$1 - SimQCDI$	% of votes
D1-M2	37.95	76
D1-H2	23.89	24

- When comparing art pieces D4-H2 versus D4-V1, the results in Table 10 were provided. This comparison was asked for similarity but also for dissimilarity checking, as a control. Hence, the 71% of the participants (67% in the inverse question, 'less' similar) answered that D4 and V1 were more similar than D4 and H2, which contrast with the similarity values obtained. Probably the contrasting colours in H2 are perceived differently by the participants than the pale colours in D4 and V1.

Table 10: Survey results and *SimQCDI* values for comparing D4-H2 versus D4-V1.

		% of votes	
	SimQCDI	similar	dissimilar
D4-H2	74.66	71	33
D4-V1	66.42	29	67

Regarding the similarities obtained between an art piece and a group of compositions by different authors, the results were the following:

- The survey asked the participants if M4 was more similar to D4-D5 or to M2-M5, and the participants' votes were summarised in Table 11. The 49% of the participants said that M4 was more similar to the D4-D5 group which is by a different author, while the 51% of the participants decided for the second group which is by the same author. The half of the participants may be influenced by the highest amount of grayish colours when relating M4 to D4-D5 (as the *SimQCDI*, see Section 11.1), while the other half may be influenced by the colour of the objects in the foreground (*red*, *blue*, *yellow* and *green*) appearing in M4 and also in M2-M5. In this case, the *SimQCDI* agreed with the opinion of the 49% of the participants.

Table 11: Survey results and *SimQCDI* values for comparing M4 to D4-D5/M2-M5 pairs.

	SimQCDI		
	M4	Average	% of votes
D4	81.39	85.41	49
D5	89.43		
M2	71.54	66.60	51
M5	49.48		

- The survey asked the participants if D2 was more similar to G1-G2 or to V1-V3 and the results gathered were those in Table 12. The similarity of pale colours in D2 and V1-V3 was obvious for 90% of the participants in the survey, while 10% found that D2 was more similar to G1-G2. In this case, the high similarity in colours between the art pieces D2 and V3, also reflected by the similarity value obtained (*SimQCDI* $= 87.83$) made the 90% of participants select the group of art compositions by Velázquez as more similar to the second (D2) art piece by Dalí, although the art pieces by Greco obtain a highest *SimQCDI* value in average. As it can be seen, a high similarity between a pair of art compositions, can condition the criterium of the participants for classifying into groups.

Table 12: Survey results and *SimQCDI* values for comparing D2 to G1-G2/V1-V3 pairs.

	SimQCDI		
	D2	Average	% of votes
G1	78.88	78.98	10
G2	79.09		
V1	65.34	76.58	90
V3	87.83		

11. Discussion

In this section, first the results obtained by the *SimQCDI* and the survey results are discussed. Then, the cognitive adequacy of the QCD model is explained relating it to the literature and to the classical colour models. In order to show specifically the contribution of the colour naming model and the similarity obtained (*SimQCD*), they are both compared to other works in the literature. Finally, in order to study if the image similarity obtained by the QCD model (*SimQCDI*) is more intuitive or consistent with human perception than standard colour-based image descriptors such as RGB histograms, a comparative is carried out.

11.1. Cognitive Adequacy of the SimQCDI measure

According to the experimentation results and the results obtained from the survey test, the hypotheses formulated in Section 10.2 can be answered:

- H1 The *SimQCDI* can be used to determine differences of art compositions belonging to the same painter (same category). However, from the survey results, it was observed that participants found hard to determine which pair of art compositions were more similar between each other if they were by the same author.

- H2 The *SimQCDI* cannot be used to differentiate between paintings of different painters. However, it can be used to identify colour similarities across painters which are obvious when noticed, but not easily seen at a first sight. From the survey results, it was observed that some participants performed better than *SimQCDI* when identifying pictures by the same authors if those art compositions contained similar objects, maybe because participants can identify shapes and spatial locations of objects, whereas *SimQCDI* is only based on colours.

  In order to find out whether human beings can be influenced or not by shapes when assigning similarities, some participants were asked to evaluate the similarity of the art compositions by *Tidying up Art*[18][42] (see an example in Figure 9). Most of them categorized qualitatively the similarity between image pairs as *quite similar*, but not as *equal*. However, it is obvious that the compositions compared have the same objects but arranged differently, therefore the colour and quantity of colours are the same in both pairs (the original and the tidied one) and the similarities provided by *SimQCDI* are 100%. As a result, it is deduced that human beings cannot abstract the colours of an art composition without being influenced by the shapes and the spatial arrangement of the objects identified in the composition. Therefore, it is concluded that *SimQCDI* and other colour indexing schemes in the literature can be useful to obtain similarities not perceptual by human beings at first sight.

- H3 Human beings can easily abstract 3D vision from 2D images and distinguish the background from the foreground in art compositions. Objects and colours in the foreground are given more importance than those in the background. It can be assumed that this fact affected the similarities assigned by the participants on the survey. This has the viceversa effect on *SimQCDI*, which cannot differentiate automatically the colours of the background

[18]Tidying up Art: `http://www.ursuswehrli.com/en/`

Table 13: Applying *SimQCDI* to art compositions D4-D5-M2-M4-M5 without background.

	D5	M2	M4	M5
D4	83.81	72.71	75.42	69.72
D5		82.43	88.07	83.18
M2			86.50	89.79
M4				88.21

from those on the foreground and therefore it is affected by the percentage of the most popular colour in the paintings.

In order to find out the adequacy of *SimQCDI* to discriminate art compositions without taking into account the background, the following proof-of-concept has been carried out on the art compositions in Table 13. The *SimQCDI* has been calculated after extracting the background colour from the histogram and normalising it.

Table 14 show the results obtained of this proof-of-concept, where it can be seen that the *SimQCDI* obtained between the art compositions is higher when the background is not considered, in the same way as the participants of the survey could automatically do. However, it is still a challenge to distinguish pixels from the background from those in the foreground while computing on-the-fly. Even sometimes human vision can also fail in distinguishing the foreground from the background in some art compositions such as H2 and H3 in Figure 8. As future work, we intend to approximate *SimQCDI* to human perception, using a learning method, but it is not the scope of the current paper.

Table 14: *SimQCDI* results using images with and without background.

	SimQCDI with background		Survey %	*SimQCDI* without background	
	M4	Average	of votes	M4	Average
D4	81.39	85.41	49	75.42	81.75
D5	89.43			88.07	
M2	71.54	66.60	51	86.50	87.35
M5	49.48			88.21	

11.2. Cognitive Adequacy of the QCD Model

According to Clark [8], the most suitable colour space to describe colour names cognitively is Hue, Saturation and Luminance (HSL), which is used by the QCD model. Furthermore, the research by Conway [3] on natural language colour-naming showed that, although it may be more accurate, people tend not to describe a colour as *dark pale blue* and may even consider this a contradiction. The same work recommended that, in order to produce more cognitive colour name descriptions, no more than one adjective should be applied to a basic colour name and also, if luminance and saturation modifiers appear equally applicable to a particular colour, the saturation modifier should be chosen. This aspect is reflected in the QCD model.

According to the studies and analysis by Kay and Regier [43, 44, 45], colour perception is language based. And, from the point of view of colour-naming research, they found and review empirically data which explain that: (i) colour categories appear to be organised around the universal colour foci, but (ii) naming differences across languages cause variations in colour cognition because colour categories are determined by language at their boundaries. Jameson and d'Andrade [46] suggested a theory supporting both tendencies: (i) colour naming may be attributed to the shape of perceptual colour space, that is, hue interacts with saturation and luminance and produce several large changes coinciding with the colour foci (*black*, *white*, *red*, *green*, *yellow* and *blue*); combined with (ii) general human cognitive tendencies toward constructing/using the most efficient name/information about colour in their society. The QCD model also combines both tendencies because it can be parameterised for describing universal colour foci or for describing specific colours which are particular from a society and also the limits of the intervals of the reference system can be adapted to the boundaries existing in each different language.

According to Palmer [1], human beings have a relational structure of colours in the mind that allows them to perceive *grey* as being intermediate between *white* and *black*. The similarity measure defined for comparing two colours in the QCD model takes into account this relational structure or colour conceptual neighbourhood.

Analysing the QCD model from the point of view of the relational structure of colours and colour complementaries, it is worth noting that there are some classical theories in the literature that explain conceptual colour oppositions. For example, Goethe's traditional colour model [36] opposed *white* $\leftrightarrow$ *black*, *red* $\leftrightarrow$ *green*, *yellow* $\leftrightarrow$ *purple* and *orange* $\leftrightarrow$ *blue* (see Section 6 in Griffin [29]), whereas Hering's traditional colour model [47] opposed *white* $\leftrightarrow$ *black*, *red* $\leftrightarrow$ *green* (like Goethe's), *yellow* $\leftrightarrow$ *blue* and *pink* $\leftrightarrow$ *brown* (see Section 6 in Griffin's paper [29] for details). Other more recent studies by Griffin (see Figure 1 in [28, 29] for more detail) showed the following oppositions: *white* $\leftrightarrow$ *black*, *yellow* $\leftrightarrow$ *purple* (like Goethe's), *red* $\leftrightarrow$ *orange*, *blue* $\leftrightarrow$ *green*, *pink* $\leftrightarrow$ *brown* (like Hering's). Finally, the opposites/complementaries for the QCD model are: *white* $\leftrightarrow$ *black* (like in Goethe's), *red* $\leftrightarrow$ *turquoise*, *orange* $\leftrightarrow$ *blue* (like in Goethe's), *yellow* $\leftrightarrow$ *purple* (like in Goethe's), *green* $\leftrightarrow$ *pink*, and the same for *pale-* and *light-* colours as it is shown in Figures 7. As far as we are concerned, there are no universal opposites for colours except for *white* $\leftrightarrow$ *black*. It seems that according to the colour space used and the population involved, the results can vary from one study to another. Moreover, the studies that have been found were usually conducted with, at the most, the 11 Basic Colour Terms (BCT) found by Berlin and Kay [41]. Possibly, by increasing the variability of colour-naming, more opposites could be found.

However, leaving the aspect of colour opposites aside, in general, the relational structures of colours by Goethe, Hering and Griffin are similar to the HSL colour space (see Section 6 in Griffin's paper [29] for details) and they are also similar to the CND obtained for the QCD model. The QCD model and its corresponding CND are completely adaptable, as it can be added more colours and assigned different weights to connections in order to reflect the social and cultural context of application.

Agent-based simulations have been carried out in the literature [48, 49, 50, 51, 52] for studying the social process of communication about colour, i.e. Komarova *et al.* [48] found that, given certain simple assumptions, a population of agents communicating about colour will converge to a system of near-optimal colour categories. Regarding these research studies, it is worth nothing that the QCD model provides a computationally adaptable way which may enable intelligent agents (i.e. robot, ambient intelligent system, web searcher, etc.): (i) to communicate among them or to a human user in a universal way (i.e. using basic colour terms) or in a specific way (i.e. using hue combinations or variations in saturation and lightness) for a particular society that understands colour names differently; and also (ii) to figure out how similar or perceptually closed are two colour names.

11.3. Comparing the QCD model and the SimQCD to Related Work

The approach presented in this paper obtains a colour model and a similarity value between colour names taking into account the spatial relational structure of the colour model selected. To the best of our knowledge, there are no works in the literature with explore the conceptual neighbourhood of colour spaces for defining similarity values.

Other works in the literature studied related topics from another perspective, i.e. colour naming and the design of a colour naming metric [16, 40]. Hence, a comparative of methodologies is carried out in this section to clarify the contribution of this paper.

With respect to colour naming, there is a great difference between the 37 colour names defined in this paper, the 267 colour names defined by Mojsilovic [16] and the 153 colour names defined by Heer and Stone [40]. This difference in the amount of colour names among colour models is given because Mojsilovic [16] and Heer and Stone [40], added new colour names to the model every time they carried out new experiments to new participants. In contrast, an objective of the approach in this paper is to find out a consensus for the majority of participants in order to not exceed the amount of colours people generally use to manage in their daily living. From the computational point of view, Mojsilovic [16] and Heer and Stone [40] presented a higher granularity in colour naming, than the QCD model which tries to approach Conway's studies [3] which declare that the basic colours that can be named by people are limited to about 10-20.

With respect to the colour naming metric, a distance is defined between colours from a cosine function by Heer and Stone [40]. The main drawback of this distance is that it only distinguishes between neighbouring colours. For the rest of the non-neighbouring colours, the given distance is the maximum (1.0), therefore the discrimination between colour names is poorer than that provided by the *SimQCD*. Table 15 (obtained from their original paper) shows the difference in the distances provided by Heer and Stone [40] and the dissimilarities provided by *SimQCD* (grey cells) which assign different dissimilarities to all colour names that allow their distinction.

Table 15: Distance between colours by Heer and Stone [40].

	Red	*Pink*	*Blue*	*Green*	*Yellow*
Red	0	0.99	1.00	1.00	1.00
Pink	0.21	0	1.00	1.00	1.00
Blue	0.64	0.42	0	1.00	1.00
Green	0.64	0.85	0.42	0	0.70
Yellow	0.42	0.64	0.64	0.21	0

With respect to the colour naming metric, the work by Mojsilovic [16] defined a distance based on the geometric property of the HSL system, where (H,S,L) are the components of the HSL colour system, which holds:

$$\Delta S = 1, \Delta H = \Delta L = 0 \rightarrow \Delta Distance = 1$$

$$\Delta L = 1, \Delta H = \Delta S = 0 \rightarrow \Delta Distance = 1$$

$$\Delta H = 1, \Delta S = \Delta L = 0 \rightarrow \Delta Distance = S\sqrt{2 \cdot (1 - \cos(1))}$$

Thus, when the saturation component is incremented in 1 unit, the distance is also incremented in 1. The same happens for lightness. Therefore, the same significance is given to a change in saturation than to a change in lightness components, whereas the *SimQCD* colour model can be tuned to give more importance to the changes in colour saturation which determine the limit between between grey colours and rainbow colours. Moreover, the distance defined by Mojsilovic [16] is not normalised, therefore a distance of 24 units obtained when calculating the similarity between two similar *red* colours cannot be assigned a high or low significance, in contrast, the *SimQCD* presented in this paper is normalised.

11.4. Comparing SimQCDI with RGB Colour Histogram Similarity

In order to evaluate if the similarity defined on the QCD model is more intuitive or consistent with human perception than standard colour-based image descriptors such as RGB histograms, a comparative is carried out in this section.

Figure 10 presents an art composition and its corresponding QCD and RGB histograms. It can be observed that the QCD histogram is more intuitive than the RGB continuous histogram since the hue and amount of colours appearing in the art composition and appearing in the QCD histogram are the same, but visualised differently, while some hues appearing in the continuous RGB histogram do not correspond to the art composition and are not so intuitive to interpret. Therefore, the RGB colour space is far from being perceptually uniform. Thus, to calculate a RGB histogram-based image similarity, it is important to obtain a good colour representation of the image by uniformly sampling the RGB space. Then, the standard 216 RGB colour palette can be used [53], and it has been the one selected in this comparison.

The quantised RGB histogram (Figure 10 (c)) is more similar to the QCD histogram (Figure 10 (a)). However, the advantage of the QCD histogram is that the colour name (semantic information) about which colour is appearing in the image is obtained, whereas the quantised RGB histogram need further interpretation of the groups of colours obtained.

For each art composition in Figure 8, the quantised RGB colour histograms has been obtained and the Euclidean distance between these RGB histograms has been computed [53] and normalised (see the Appendix), which is denoted by *SimRGB*. Then, *SimRGB* and *SimQCDI* are compared in order to analyse which of these methods is closer to the results of the survey described previously in Section 10.4:

- When comparing the art pieces D1-D4-D5, the results obtained are shown in Table 16. Considering that the most cognitive result is the opinion of the participants in the survey, then the order of voting results, which is (23, 37, 40), is important, and the similarities obtained should follow this order and have a similar quantisation to be intuitive/cognitive enough. The *SimQCDI* obtains the following values (76.83, 79.50, 77.91), which involves that two of the values (79.50 and 77.91) must change the position to get the most cognitive order. The *SimRGB* provides the values (81, 87, 77.91) which needs two changes to get to the order of the responses of the participants (first (87 and 77.91) and after (81 and 77.91)). Hence, the *SimQCDI* is more coherent with the participants of the survey.

- When comparing the art pieces G1-G2-G3, the results obtained are those in Table 17. Considering the opinion of the participants surveyed, the most intuitive order of similarity

Table 16: Survey answers and *SimQCDI* versus *SimRGB* results for images D1-D4-D5.

	SimQCDI		% of votes		*SimRGB*	
	D4	D5	D4	D5	D4	D5
D1	76.83	77.91	23	40	81.0	79.0
D4	–	79.50	–	37	–	87.0

is (16, 17, 70). The values provided by *SimQCDI* and *SimRGB* are (80.93, 80.55, 84.70) and (91.0, 87,0, 86.0), respectively. In this case, the *SimQCDI* agrees completely with the opinion of the survey, since the difference between 16 and 17 is very small, such as the difference between 80.93 and 80.55. However, *SimRGB* is far from the correct order since the values provided differ greatly both in order and value.

Table 17: Survey answers and *SimQCDI* versus *SimRGB* results for images G1-G2-G3.

	SimQCDI		% of votes		*SimRGB*	
	G2	G3	G2	G3	G2	G3
G1	84.70	80.93	67	16	86.0	91.0
G2	–	80.55	–	17	–	87.0

- When comparing the art pieces H1-H2-H4, the similarities and votes gathered are shown by Table 18. In this situation, *SimQCDI* and *SimRGB* have similar performance, since they follow the order provided by the survey, (22, 32, 42), with values of *SimQCDI* = (82.86, 82.90, 89.28), and *SimRGB*=(87, 88, 92). They both agree with the participants of the voting, in the same order but a bit far from the opinion of the participants in the survey.

Table 18: Survey answers and *SimQCDI* versus *SimRGB* results for images H1-H2-H4.

	SimQCDI		% of votes		*SimRGB*	
	H2	H4	H2	H4	H2	H4
H1	89.28	82.86	46	22	92.0	87.0
H2	–	82.90	–	32	–	88.0

- When comparing the art pieces V1-G2 versus V1-D4, the results were those in Table 19. In this situation, *SimQCDI* and *SimRGB* have similar performance.
- When comparing the art pieces D1-M2 versus D1-H2, the results obtained are those in Table 20. In this situation, *SimQCDI* and *SimRGB* have similar performance: same order and same quantisation.
- When comparing art pieces D4-H2 versus D4-V1, the similarities and votes are those in Table 21. In this case, *SimQCDI* and *SimRGB* have similar performance.

Table 19: Survey answers and *SimQCDI* versus *SimRGB* results for V1-G2/V1-D4 pairs.

	SimQCDI	% of votes	*SimRGB*
V1-G2	61.83	50	65.0
V1-D4	66.42	50	70.0

Table 20: Survey answers and *SimQCDI* versus *SimRGB* results for D1-M2/D1-H2 pairs.

	$1 - SimQCDI$	% of votes	*SimRGB*
D1-M2	37.95	76	37.0
D1-H2	23.89	24	19.0

Table 21: Survey answers and *SimQCDI* versus *SimRGB* results for D4-H2/D4-V1 pairs.

		% of votes		
	SimQCDI	similar	dissimilar	*SimRGB*
D4-H2	74.66	71	33	80.0
D4-V1	66.42	29	67	70.0

- When the survey asked the participants if M4 was more similar to D4-D5 or to M2-M5, and the results obtained are summarised in Table 22. In this case, *SimQCDI* and *SimRGB* perform similarly. However the difference between the average of *SimRGB* is higher (23 points) than the difference between the averages of *SimQCDI* (18.21 points), while the participants voting is distributed approximately at 50%. *SimQCDI* finds out less differences in colour than *SimRGB*, as the participants do.

Table 22: Survey answers and *SimQCDI* versus *SimRGB* results for M4 with respect to D4-D5/M2-M5.

	SimQCDI		%	*SimRGB*	
	M4	Average	of votes	M4	Average
D4	81.39	85.41	49	88.0	88.0
D5	89.43			88.0	
M2	71.54	66.60	51	62.0	65.0
M5	49.48			68.0	

- When the survey asked the participants if D2 was more similar to G1-G2 or to V1-V3, the results obtained are shown by Table 23. In this case, *SimQCDI* and *SimRGB* disagree with the opinion of the participants, but the difference in values obtained by *SimRGB* is 5.5 points, while the difference by *SimQCDI* is 2.4 points.

Table 23: Survey answers and *SimQCDI* versus *SimRGB* results for D2 with respect to G1-G2/V1-V3.

<table>
<tr><th></th><th colspan="2">SimQCDI</th><th>%</th><th colspan="2">SimRGB</th></tr>
<tr><th></th><th>D2</th><th>Average</th><th>of votes</th><th>D2</th><th>Average</th></tr>
<tr><td>G1</td><td>78.88</td><td rowspan="2">78.98</td><td rowspan="2">10</td><td>81.0</td><td rowspan="2">81.5</td></tr>
<tr><td>G2</td><td>79.09</td><td>82.0</td></tr>
<tr><td>V1</td><td>65.34</td><td rowspan="2">76.58</td><td rowspan="2">90</td><td>67.0</td><td rowspan="2">76.0</td></tr>
<tr><td>V3</td><td>87.83</td><td>85.0</td></tr>
</table>

In summary, after comparing *SimQCDI* and *SimRGB* the main conclusions are:

- *SimQCDI* is more intuitive; and
- although the quantisation of both *SimQCDI* and *SimRGB* can be considered as equivalent, the comparative results obtained from the survey show that *SimQCDI* is slightly more adequate than *SimRGB*.

12. Conclusions and Future Work

A model for Qualitative Colour Description (QCD) based on HSL colour space has been presented and proved to name colours in a general and adaptive way by distinguishing *rainbow* colours, *pale*, *light*, and *dark* colours and colours in the *grey* scale. The relational structure of the QCD model is also analyzed by means of a conceptual neighbourhood diagram.

A measure of similarity between colour names has also been defined taking into account the relational structure in QCD (*SimQCD*). *SimQCD* is unique and showed to fulfill interesting and intuitive properties to solve absolute and relative comparisons of qualitative colours.

Furthermore, a similarity measure between colour images (*SimQCDI*) has been presented and proved: (i) to determine colour differences of art compositions belonging to the same painter; (ii) to identify colour similarity across authors; and (iii) to agree with most of the results of a survey test on these art compositions carried to participants. From the results, we conclude that, only by using the *SimQCDI*, we cannot determine if two art compositions were painted by the same artist or not. This could be achieved by studying the authors' palette and formulating a classification algorithm which make use of learning techniques (i.e support vector machine, neural network, etc.). This research work is intended to be carried out by the authors in the future.

The differences between the results of the survey test and the results of the *SimQCDI* approach drove us to carry out two proofs-of-concept to investigate whether: (a) human beings cannot discard shapes/objects' locations when comparing art compositions and (b) their ability to abstract the foreground from the background when assigning similarities. These proofs-of-concept confirmed those skills which contrasted with the *SimQCDI* approach which only considers colour palettes. However, those proofs also concluded that *SimQCDI* as other colour indexing schemes can provide colour similarities across painters which are not perceptual for participants at a first sight.

The cognitive adequacy of the QCD model has also been proved from the point of view of colour naming in natural language and from the point of view of the relational structures of colour perception in classical theories and in psychological studies. Moreover, the *SimQCD* measure has been compared to other works in the literature. Finally, in order to study if the image similarity defined by the QCD model (*SimQCDI*) is more intuitive or consistent with human perception than standard colour-based image descriptors such as RGB histograms, a comparison is done.

As future work, we intend to improve the *SimQCDI* similarity measure in order to reflect cognitive aspects found in the survey, such as: (i) avoiding the background in the comparisons; (ii) taking into account colour contrasts when comparing colour compositions; (iii) extending the similarity measure to include the shape and location of the objects in the art composition; and (iv) applying a learning algorithm in order to classify the art compositions by authors. Another important topic to study is the influence of the weights used in the *SimQCD* model which were parameterised using values as a baseline in this paper.

Acknowledgments

This work was supported by European Commission through FP7 Marie Curie IEF actions under project COGNITIVE-AMI (GA 328763), the interdisciplinary Trasregional Collaborative Research Center Spatial Cognition SFB/TR 8, the Deutscher Akademischer Austausch Dienst (DAAD), Andalusian Regional Ministry of Economy (project SIMON TIc-8052), Spanish Ministry of Economy and Competitiveness (project TIN2011-24147), Generalitat Valenciana (project GVA/2013/135) and Universitat Jaume I (project P11B2013-29).

We acknowledge the reviewers for their comments which help us to improve this paper. We also acknowledge that the paintings by Velázquez and Greco are on public domain under creative commons copyrights. We thank the VG Bild-Kunst in Bonn, Germany, for providing us 'free fair use' copyrights for the paintings by Miró and Dalí under scientific quotation. We also thank Hundertwasser foundation in Vienna, Austria, for providing us 'free fair use' copyrights for the works by Hundertwasser under scientific quotation. We thank Kein & Aber AG to allow us to include a painting by Wehrli in our paper.

References

[1] S. Palmer, Vision Science: Photons to Phenomenology, MIT Press, 1999.

[2] A. Chapanis, Color names for colour space, American Scientist 53 (3) (1965) pp. 327–346. URL `http://www.jstor.org/stable/27836112`

[3] D. Conway, An experimental comparison of three natural language colour naming models, in: Proc. East-West International Conference on Human-Computer Interactions, 1992, pp. 328–339.

[4] T. Richter, R. Zwaan, Processing of colour words activates colour representations, Cognition 111 (2009) 383–389.

[5] L. Connell, Representing object colour in language comprehension, Cognition 102 (2007) 476–485.

[6] Z. Falomir, E. Jiménez-Ruiz, M. T. Escrig, L. Museros, Describing Images using Qualitative Models and Description Logics, Spatial Cognition and Computation, 11 (1) (2011) 45–74.

[7] D. J. Therriault, On the future of object recognition: The contribution of color, in: T. P. Cao (Ed.), Object Recognition, InTech, 2011, pp. 3–14.

[8] A. Clark, Perception: Colour, in: W. Bechtel, G. Graham (Eds.), A Companion to Cognitive Science, Blackwell Companions to Philosophy, Blackwell Publishers Inc, Malden, MA, USA, 1998, pp. 282–288.

[9] D. Nickerson, History of the Munsell color system, company, and foundation, COLOR Research and Application 1 (1) (1976) 7–10.

[10] M. Fairchild, Color appearance models, Wiley-IS&T series in imaging science and technology, J. Wiley, 2005.

[11] G. Wyszecki, W. S. Stiles, Color Science: concepts and methods, quantitative data and formulae, New York: Wiley, 1982.

[12] N. Moroney, M. D. Fairchild, R. W. G. Hunt, C. Li, M. R. Luo, T. Newman, The CIECAM02 color appearance model, in: IST/SID 10 th Color Imaging Conference: Color Science, System and Application, 2002, pp. 23–27.

[13] M. Sarifuddin, R. Missaoui, A new perceptually uniform color space with associated color similarity measure for content based image and video retrieval, in: Multimedia Information Retrieval Workshop, 28th annual ACM SIGIR conference, 2005, pp. 3–7.

[14] J. Van De Weijer, C. Schmid, Applying color names to image description., in: ICIP (3), IEEE, 2007, pp. 493–496. URL `http://dblp.uni-trier.de/db/conf/icip/icip2007-3.html#WeijerS07`

[15] G. Menegaz, A. L. Troter, J. Sequeira, J. M. Boi, A discrete model for color naming, EURASIP J. Appl. Signal Process, Special Issue on Image Perception (2007) 1–10.

[16] A. Mojsilovic, A computational model for color naming and describing color composition of images, IEEE Transactions on Image Processing 14 (5) (2005) 690–699.

[17] M. Seaborn, L. Hepplewhite, T. J. Stonham, Fuzzy colour category map for the measurement of colour similarity and dissimilarity, Pattern Recognition 38 (2) (2005) 165–177.

[18] E. Eriki, C. Yener, N. Camgöz, Colour naming, Optics and Laser Technology 38 (2006) 466–485.

[19] Y. Liu, D. Zhang, G. Lu, W. Ma, Region-based image retrieval with perceptual colors, in: Proc. of Pacific-Rim Multimedia Conference (PCM2004), 2004, pp. 931–938.

[20] J. M. Corridoni, A. D. Bimbo, E. Vicario, Image retrieval by color semantics with incomplete knowledge, Journal of the American Society for Information Science 49 (3) (1998) 267–282.

[21] P. L. Stanchev, D. G. Jr., B. Dimitrov, High level colour similarity retrieval, International Journal on Information Theories and Applications 10 (3) (2003) 363–369.

[22] J. M. G. Lammens, A computational model of color perception and color naming, Ph.D. thesis, Faculty of the Graduate School of State University of New York at Buffalo, USA (1994).

[23] T. Berk, L. Brownston, A. Kaufman, A new color-naming system for graphics languages, IEEE Computer Graphics and Applications 2 (1982) 37–44.

[24] M. Vik, Industrial colour difference evaluation: LCAM textile data, in: AIC 2004 Color and Paints, Interim Meeting of the International Color Association, Proceedings, 2004, pp. 138–142.

[25] K. N. Plataniotis, A. N. Venetsanopoulos, Color image processing and applications, Springer-Verlag, NY, USA, 2000.

[26] Z. Falomir, L. Museros, L. Gonzalez-Abril, I. Sanz, A model for qualitative colour comparison using interval distances, Displays 34 (2013) 250–257.

[27] M. R. Luo, G. Cui, B. Rigg, The development of the CIE 2000 colour-difference formula: CIEDE2000, COLOR Research and Application 26 (2001) 340–350.

[28] L. D. Griffin, Similarity of psychological and physical colour space shown by symmetry analysis, COLOR Research and Application 26 (2) (2001) 151–157.

[29] L. D. Griffin, Basic colour foci and landmarks of the body colour solid, Journal of Vision 6 (13) (2006) 47–47.

[30] G. H. Joblove and D. Greenberg. Color spaces for computer graphics. SIGGRAPH Comput. Graph. 12, 3 (1978), 20–25.

[31] C. Ware. Information Visualization. Perception for Design. Elsevier, 2012.

[32] L. Steels, T. Belpaeme. Coordinating perceptually grounded categories through language: a case study for colour. Behav. Brain. Sci. (4): 469 – 489, discussion 489 – 529, 2005.

[33] L. Gonzalez-Abril, F. Cuberos, F. Velasco, J. Ortega, Ameva: An autonomous discretization algorithm, Expert Systems with Applications 36 (3, Part 1) (2009) 5327 – 5332.

[34] C. Freksa, Conceptual neighborhood and its role in temporal and spatial reasoning, in: M. G. Singh, L. Travé-Massuyès (Eds.), Proceedings of the IMACS Workshop on Decision Support Systems and Qualitative Reasoning, North-Holland, Amsterdam, 1991, pp. 181–187.

[35] C. Kennard, M. Lawden, A. B. Morland and K. H. Ruddock, Colour Identification and Colour Constancy are Impaired in a Patient with Incomplete Achromatopsia Associated with Prestriate Cortical Lesions, Proc R Soc London Ser B 260:169 – 175,1995.

[36] J. W. v. Goethe, Goethe's color theory / arranged and edited by Rupprecht Matthaei, english ed. / translated and edited by Herb Aach. Edition, Studio Vista, London, 1971.

[37] L. González, C. Angulo, F. Velasco, A. Català, Dual unification of bi-class support vector machine formulations, Pattern Recognition 39 (7) (2006) 1325–1332.

[38] R. Roja, Neural Networks: A systematic introduction, Springer-Verlag, 1996.

[39] J. Quinlan, C 4.5 Programs for Machine Learning, Morgan Kaufmann, 1993.

[40] J. Heer, M. Stone, Color Naming Models for Color Selection, Image Editing and Palette Design, ACM Human Factors in Computing Systems (CHI), pp.1007–1016, 2012.

[41] B. Berlin, P. Kay, Basic color terms: Their universality and evolution., University of California Press, Berkeley, 1969.

[42] U. Wehrli, Tidying up Art, Kein and Aber, 2003.

[43] P. Kay, T. Regier, Colour naming universals: The case of berinmo, Cognition 111 (2007) 289–298.

[44] P. Kay, T. Regier, Language, thought, and color: Recent developments, Trends in Cognitive Sciences 10 (2) (2006)

51–54.

[45] T. Regier, P. Kay, Language, thought, and colour: Whorf was half right, Trends in Cognitive Sciences 13 (10) (2009) 439–446.

[46] K. Jameson, R. D'Andrade, It's not really red, green, yellow, blue: An inquiry into cognitive color space, in: C. Hardin, L. Maffi (Eds.), Color categories in thought and language, Cambridge University Press, Cambridge, 1997, pp. 295–319.

[47] C. Baumann, Ewald Hering's opponent colors. History of an idea, Ophthalmologe 89 (3) (1992) 249–52.

[48] N. Komarova, K. Jameson, L. Narens, Evolutionary models of color categorization based on discrimination, Journal of Mathematical Psychology 51 (2007) 359–382.

[49] N. Komarova, K. Jameson, Population heterogeneity and color stimulus heterogeneity in agent-based color categorization, Journal of Theoretical Biology 253 (2008) 680–700.

[50] K. Jameson, N. Komarova, Evolutionary models of color categorization. I. Population categorization systems based on normal and dichromat observers, Journal of the Optical Society of America 26 (2009) 1414–1423.

[51] K. Jameson, N. Komarova, Evolutionary models of color categorization. II. Realistic observer models and population heterogeneity, Journal of the Optical Society of America 26 (2009) 1424–1436.

[52] T. Griffiths, M. Kalish, Language evolutiion by iterated learning with bayesian agents, Cognitive Science 31 (2007) 441–480.

[53] K. Meskaldji, S. Boucherkha, S. Chikhi, Color Quantization and its Impact on Color Histogram Based Image Retrieval, CEUR Wokshop Proc. of the 2nd Conférence Internationale sur l'Informatique et ses Applications (CIIA'09), vol. 547, Saida, Algeria, May 3-4, 2009.

Appendix

The RGB colour histograms of each art composition in Figure 8 are obtained and quantised to 216 colours. The Euclidean distance between these RGB histograms has been computed [53]:

$$d^2(h_1,h_2) = \sum_r \sum_g \sum_b (h_1(r,g,b) - h_2(r,g,b))^2$$

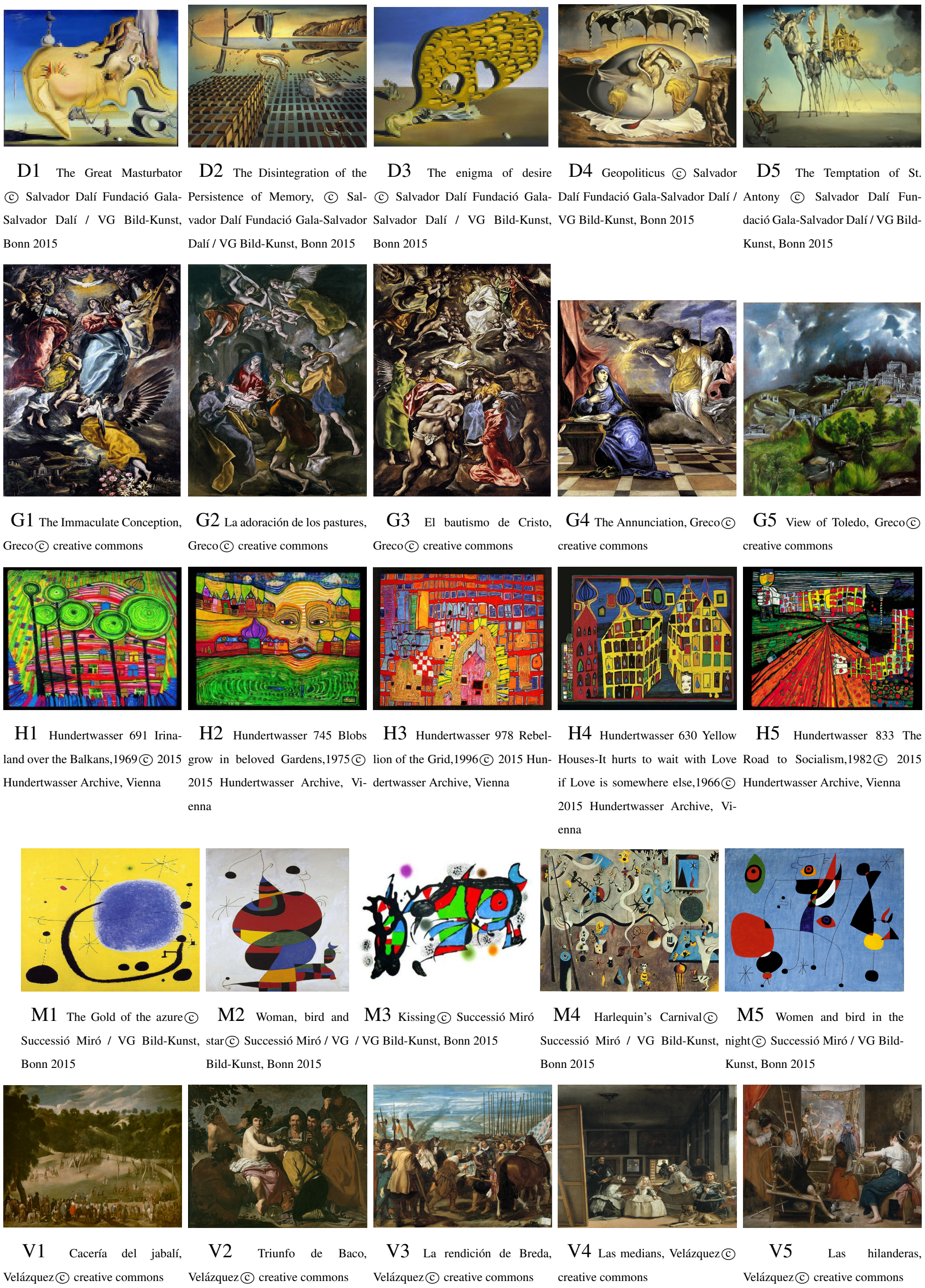
This distance has been also normalised to get values between 0 and 100:

$$SimRGB = 100 \cdot (1 - \frac{d}{MaxDistance})$$

The results obtained are shown in Table 24.

Table 24: Euclidean distance applied to RGB histograms quantised to 216 colours obtained for each art composition in Figure 8.

	D2	D3	D4	D5	G1	G2	G3	G4	G5	H1	H2	H3	H4	H5	M1	M2	M3	M4	M5	V1	V2	V3	V4	V5
D1	78.0	77.0	81.0	79.0	76.0	72.0	76.0	81.0	80.0	82.0	81.0	83.0	79.0	75.0	59.0	63.0	57.0	82.0	69.0	63.0	62.0	75.0	61.0	68.0
D2		78.0	89.0	84.0	81.0	82.0	82.0	84.0	84.0	78.0	76.0	78.0	78.0	73.0	56.0	58.0	55.0	84.0	65.0	67.0	74.0	85.0	73.0	82.0
D3			80.0	77.0	74.0	74.0	76.0	79.0	81.0	79.0	78.0	79.0	77.0	72.0	58.0	61.0	55.0	80.0	69.0	69.0	66.0	78.0	65.0	72.0
D4				87.0	83.0	82.0	85.0	89.0	85.0	81.0	80.0	82.0	82.0	77.0	58.0	60.0	56.0	88.0	67.0	70.0	72.0	83.0	70.0	78.0
D5					77.0	76.0	78.0	82.0	80.0	77.0	75.0	77.0	76.0	72.0	56.0	58.0	55.0	88.0	65.0	64.0	67.0	80.0	68.0	74.0
G1						86.0	91.0	88.0	84.0	81.0	81.0	82.0	86.0	84.0	56.0	59.0	56.0	80.0	68.0	61.0	71.0	82.0	71.0	78.0
G2							87.0	81.0	83.0	76.0	75.0	77.0	80.0	76.0	54.0	58.0	54.0	76.0	64.0	65.0	82.0	86.0	79.0	86.0
G3								88.0	84.0	81.0	81.0	82.0	87.0	83.0	56.0	59.0	56.0	80.0	67.0	64.0	74.0	83.0	72.0	79.0
G4									87.0	86.0	85.0	86.0	87.0	84.0	59.0	61.0	58.0	87.0	71.0	66.0	67.0	81.0	66.0	75.0
G5										85.0	83.0	85.0	85.0	81.0	59.0	61.0	58.0	84.0	72.0	70.0	70.0	83.0	68.0	77.0
H1											92.0	91.0	87.0	85.0	61.0	62.0	59.0	82.0	73.0	65.0	63.0	75.0	61.0	69.0
H2												89.0	88.0	86.0	60.0	61.0	58.0	80.0	72.0	64.0	61.0	73.0	59.0	67.0
H3													87.0	84.0	62.0	62.0	59.0	83.0	74.0	65.0	64.0	76.0	62.0	70.0
H4														87.0	61.0	61.0	58.0	80.0	71.0	65.0	66.0	77.0	64.0	72.0
H5															57.0	58.0	57.0	76.0	70.0	61.0	60.0	71.0	59.0	66.0
M1																47.0	45.0	59.0	57.0	48.0	46.0	54.0	45.0	51.0
M2																	47.0	62.0	54.0	50.0	50.0	60.0	50.0	55.0
M3																		58.0	52.0	47.0	47.0	54.0	46.0	51.0
M4																			68.0	67.0	66.0	81.0	65.0	74.0
M5																				56.0	53.0	64.0	52.0	59.0
V1																					61.0	63.0	53.0	62.0
V2																						81.0	87.0	87.0
V3																							81.0	89.0
V4																								87.0



D1 The Great Masturbator © Salvador Dalí Fundació Gala-Salvador Dalí / VG Bild-Kunst, Bonn 2015

D2 The Disintegration of the Persistence of Memory, © Salvador Dalí Fundació Gala-Salvador Dalí / VG Bild-Kunst, Bonn 2015

D3 The enigma of desire © Salvador Dalí Fundació Gala-Salvador Dalí / VG Bild-Kunst, Bonn 2015

D4 Geopoliticus © Salvador Dalí Fundació Gala-Salvador Dalí / VG Bild-Kunst, Bonn 2015

D5 The Temptation of St. Antony © Salvador Dalí Fundació Gala-Salvador Dalí / VG Bild-Kunst, Bonn 2015

G1 The Immaculate Conception, Greco© creative commons

G2 La adoración de los pastures, Greco© creative commons

G3 El bautismo de Cristo, Greco© creative commons

G4 The Annunciation, Greco© creative commons

G5 View of Toledo, Greco© creative commons

H1 Hundertwasser 691 Irinaland over the Balkans,1969© 2015 Hundertwasser Archive, Vienna

H2 Hundertwasser 745 Blobs grow in beloved Gardens,1975© 2015 Hundertwasser Archive, Vienna

H3 Hundertwasser 978 Rebellion of the Grid,1996© 2015 Hundertwasser Archive, Vienna

H4 Hundertwasser 630 Yellow Houses-It hurts to wait with Love if Love is somewhere else,1966© 2015 Hundertwasser Archive, Vienna

H5 Hundertwasser 833 The Road to Socialism,1982© 2015 Hundertwasser Archive, Vienna

M1 The Gold of the azure© Successió Miró / VG Bild-Kunst, Bonn 2015

M2 Woman, bird and star© Successió Miró / VG Bild-Kunst, Bonn 2015

M3 Kissing© Successió Miró / VG Bild-Kunst, Bonn 2015

M4 Harlequin's Carnival© Successió Miró / VG Bild-Kunst, Bonn 2015

M5 Women and bird in the night© Successió Miró / VG Bild-Kunst, Bonn 2015

V1 Cacería del jabalí, Velázquez© creative commons

V2 Triunfo de Baco, Velázquez© creative commons

V3 La rendición de Breda, Velázquez© creative commons

V4 Las medians, Velázquez© creative commons

V5 Las hilanderas, Velázquez© creative commons

Figure 8: Testing Scenario, art pieces by the following painters: Dalí (D1-5), el Greco (G1-5), Hundertwasser (H1-5), Miró (M1-5) and Velázquez (V1-5).

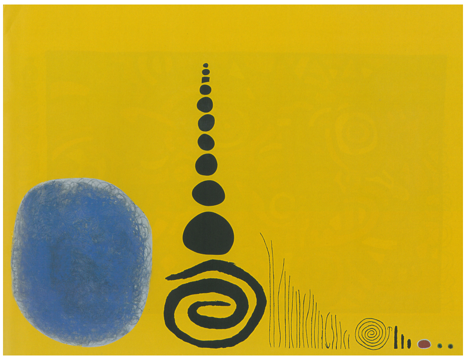
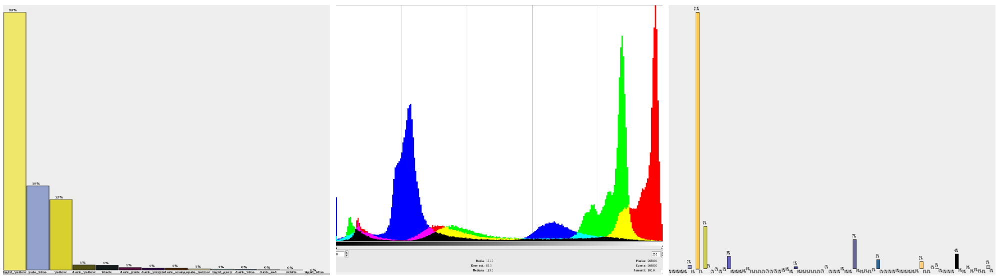


Figure 9: *Mirós L'or de l'azur aufräumen* by U. Wehrli [42] © Kein & Aber AG. Note that it is Miró's painting M1 in Figure 8, but tidied up.



(a) (b) (c)

Figure 10: Comparison of colour histograms corresponding to painting M1 in Figure 8: (a) QCD histogram, (b) continuous RGB histogram, and (c) discretised RGB histogram to 216 colours.